

Washington's Teacher and Principal Evaluation System: Goal Setting and Measures of Student Growth for Instructional Improvement

Final Report

Prepared for the Office of Superintendent of Public Instruction

Margaret L. Plecki
Ana M. Elfers
Elise St. John
Theresa Ling Yeh

University of Washington
College of Education
Educational Leadership, Policy and Organizations

January 2015

Table of Contents

Executive Summary

Introduction	1
Background on the Teacher and Principal Evaluation Project	1
Brief summary of relevant literature on teacher evaluation	3
Purposes of teacher evaluation.....	3
Approaches to design and implementation	4
Research Questions and Methods	5
Methods	6
Interview Data	7
Survey Data	7
Examining the Early Implementation of TPEP: Findings	9
Variation and Innovation in Early Implementation (Years One and Two).....	9
Goal Setting and Collection of Evidence	18
Role of Collaboration.....	28
Evolving Professional Development.....	33
Restructuring Human Resources	37
Role of Leadership	40
Use of Electronic Tools	44
Views of Educators about the Impact of the Revised Evaluation System	49
Policy Implications and Conclusions	54
Summary and Concluding Comments	54
Potential Policy Implications.....	55
References	57

Figures

Figure 1:	Teacher Survey Participants by School Level and Teaching Experience.....	8
-----------	--	---

Tables

Table 1:	Characteristics of Study Site Districts.....	6
Table 2:	2013 Survey: Teachers' Ratings of the Usefulness of Various Sources in Informing their Goal Setting for Student Growth.....	21
Table 3:	2014 Survey: Teacher Views on Specific Supports and Evaluation Issues	21
Table 4:	Types of Assessments Teachers Plan to Use as Evidence of Student Growth for Purposes of the Evaluation	23
Table 5:	Teacher Confidence Regarding Goal Setting and Evidence of Student Growth.....	25
Table 6:	Perceived Goal Setting Challenges Faced by Teachers	26
Table 7:	My District Has Provided Examples of Measures/Evidence of Student Learning that Teachers Can Use	26
Table 8:	2014 Survey: Teacher Collaboration on Activities Related to Evaluation.....	30
Table 9:	2014 Survey: Extent of Teacher Collaboration on Activities Related to Evaluation with Grade Level, Department, or PLC	31
Table 10:	2013 Survey: Teacher and Principal Views on the Collection of Evidence of Student Growth.....	32
Table 11:	2014 Survey: Teacher Views of Evaluation System.....	50
Table 12:	Perceived Impact of TPEP Implementation.....	51
Table 13:	2014 Survey: Principal Views on Teacher Evaluation System	52
Table 14:	2014 Survey: Perceived Obstacles to TPEP Implementation	53

Executive Summary

This report examines the initial implementation of Washington’s Teacher and Principal Evaluation Project (TPEP), with a specific focus on how educators set goals and use student growth measures to demonstrate achievement of these goals. Through an in-depth analysis of six districts and nineteen schools, we consider the issues Washington schools and districts face in implementing the revised evaluation system, including ways the evaluation process has prompted opportunities for teachers and other professionals to engage in professional learning for the purpose of instructional improvement, and how eVAL and other electronic tools have been used in this process.

Using a mixed-methods design that included both surveys and qualitative case studies, we follow the implementation efforts of six districts across Washington state over two years. As early implementers of TPEP through participation in pilot or Regional Implementation Grants (RIGs), the strategic selection of these districts provides a window into the issues and challenges faced by districts that have sought to authentically and constructively engage in the evaluation process.

The introduction of a change as massive as a revised statewide evaluation system has involved educators across multiple levels of the system, prompting a restructuring of human resources and evolving professional development and supports. The findings from this study suggest that the initial implementation of the revised evaluation system created some conditions for the improvement of teaching and learning, but also resulted in additional challenges for local school systems. Though the school and district contexts differed considerably from one another, the following themes illustrate aspects of TPEP implementation that were common across multiple settings.

Variation and Innovation in Early TPEP Implementation

Among all six case study districts, educational leaders directly connected the evaluation system to an instructional improvement agenda. They adopted a learning stance towards teacher evaluation, emphasizing the potential benefits rather than viewing the revised system as yet another state mandate requiring compliance. The use of an instructional framework in guiding improvement efforts provided a rich medium for conversations about effective instructional practice. Survey data from both 2013 and 2014 confirm that a majority of teachers and building administrators find the instructional frameworks adopted by their districts to be a good model for effective teaching. These educators recognize that the teacher evaluation policy has the potential to fundamentally change how teaching and learning happen, but also that it requires framing the conversation in ways that productively engage staff in support of authentic long-term professional growth.

All of the case study districts were strategic in how they approached introducing TPEP to their staff. School leaders frequently leveraged existing staff capacity to support early implementation efforts. For example, a common theme among the districts was to pilot the revised evaluation system with well-respected veteran teachers. In doing so, these districts not only developed advocates for the revised system, which helped to elicit buy-in among teachers, but they also created a cadre of teachers who could then support their colleagues.

For the initial state rollout in 2013-14, the case study districts adopted different implementation approaches for transitioning their staff to the revised evaluation system. Four of the six districts decided that virtually all teachers would be evaluated on either a comprehensive or a focused evaluation in the first year of statewide implementation. The other two districts chose to stagger staff participation over three years, due in part to a large number of teachers on provisional contracts or to negotiations with their teachers' associations. Several districts chose to incrementally introduce their staff to the instructional framework over several years, by focusing on different components of the framework and targeting the evaluation on those dimensions.

The case study districts all used familiar delivery mechanisms to provide support to teachers and principals, through early intensive phases of professional development. But several districts did not require that teachers be evaluated on all the framework components in their evaluation plan in the beginning. District leaders explained that this decision was due in part to a lack of clarity and guidance from the state on certain aspects of the evaluation, but also because they believed their staff needed more time to become familiar with the framework components as well as the overall evaluation process. Similar to the notion of districts incrementally introducing the framework and targeting the evaluation around those dimensions, some districts and schools chose to selectively focus professional learning on specific aspects of the framework, even while they evaluated teachers across all the dimensions. In these cases, educators focused on the framework dimensions that were most salient to their day-to-day professional realities.

Goal Setting and Collection of Evidence

During the first two years of statewide implementation, goal setting and the collection of evidence for student growth proved to be among the most challenging aspects of the revised evaluation system. Findings suggest that teachers struggle with setting goals for student growth and linking these growth goals to initiatives aimed at improvement, while principals find it difficult to allocate the requisite time to adequately help teachers develop measures to assess student learning. The challenge of how to practically approach goal setting does not diminish the fact that most educators involved in TPEP believe that paying attention to student growth provides some benefit to the teacher evaluation process. In the 2014 survey, three-quarters of teachers agree that examining student growth is a useful part of evaluation.

The majority of teachers use classroom-based assessments, pre/post unit tests, or assessments developed by their department, grade level or Professional Learning Communities (PLCs) to support student growth goals. School and district administrators report that they primarily encourage their teachers to use these types of assessments. Overall, the majority of school administrators do not encourage the use of state tests. Educators in the case study districts indicated a need for additional, appropriate assessments which can support the integrity and reliability of the evaluation process. In addition, they expressed a desire for training in how to use assessments to inform student growth goals.

Role of Collaboration

By design, the revised teacher and principal evaluation system was intended to broaden and deepen the conversation among educators about high quality instructional practice. Early conversations with teachers' associations proved critical to productive engagement around this issue. The case study districts sought to be intentional about including association representatives in discussions of how the implementation would play out, and the best ways to support staff throughout the process.

At the building level, collaboration around TPEP most frequently occurred between the principal and individual teachers, and in groups of teachers working together. PLCs and grade level or department teams were natural places for teachers to work together on issues related to TPEP. Yet, how well these spaces functioned and whether or not they were even used for activities related to TPEP, varied considerably by building, grade level and department.

Several activities in teacher evaluation may be associated with collaborative efforts, including goal setting, collecting evidence of student growth, designing assessments, and analyzing student growth data together with colleagues. Over half of teachers reported working on goal setting for student growth and designing classroom-based assessments mostly with others. Conversely, a majority reported collecting evidence of student growth and analyzing student growth data mostly by themselves. Variation on these items was found by district and also school level, as elementary school teachers were significantly more likely than high school teachers to work with others on goal setting, designing classroom-based assessments, and analyzing growth data, and middle school teachers were significantly more likely than high school teachers to work with others on collecting evidence.

The TPEP process appears to push professional learning by exploring what instructional practices may be lacking within school settings, and what supports might be needed for teachers. It also reveals that teachers in particular roles (e.g., special education, PE, art, and other specialists) may benefit from greater integration and inclusion in professional learning communities. Many principals and teachers affirmed the richness of conversations about instructional practice, and considered this aspect to be among the most beneficial of the entire process, though finding the time necessary for these conversations was a challenge. TPEP provided some teachers with the opportunity to be actively involved in their evaluation by contributing evidence of their teaching practices, thereby having a stronger voice in the process.

Evolving Professional Development

After the first 18 months of statewide implementation, school and district staff now have a greater understanding of the TPEP process. While the standard statewide training available to districts has been widely used, some of the case study districts are now focusing on their own locally developed and embedded training. Decisions about the content and format of professional training are being driven partially by local concerns and the specific needs of teachers. In some cases there has been a shift to working with novice teachers or staff new to the district, who may not be familiar with the district's framework.

The instructional frameworks have also identified areas where educators have gaps in their knowledge, such as assessment literacy. Formative assessment, in particular, is essential for the evaluation, and is an area where additional training may be needed. Teachers may understand the student growth requirement, but they may not know how to choose the right type of assessment to demonstrate student growth, or how to match the assessment to the needs of their students. Nearly three-quarters of teachers surveyed identified “training in assessments that can be used to determine student growth” and “strategies for collecting different forms of evidence of student learning” as topics that would be useful to them.

Building administrators carry much of the load in the implementation of TPEP, and they have needed substantial support in learning the evaluation model. All of the case study districts utilized some form of leadership development for their building administrators. This often took the form of regular weekly or monthly meetings in which TPEP was a major focus. As part of their training, but also to address concerns about rater agreement across school sites, building administrators in several districts regularly participated in “rounds” to schools to observe teaching or watch videotaped lessons, in order to discuss and try to come closer to agreement about what good instruction should look like. The issue of rater agreement and how to ensure that principals implemented TPEP consistently across schools was a concern expressed by many individuals and levels of the school system.

Restructuring Human Resources

Arguably the greatest challenges to effectively implementing TPEP are time and human resources. The workload for building administrators can be overwhelming. An equally daunting task is managing the financial costs to support the implementation of both the revised evaluation system and the Common Core State Standards (CCSS), given the need for additional staff and professional development support. All six of the case study districts restructured human resources in order to accommodate the increased workload.

Along these lines, TPEP implementation may differentially impact elementary and secondary school administrators. It has been suggested that TPEP implementation is more challenging at the elementary level because of the workload on the principal. However, differential impact on administrators is not always related to the building level. The experience, management or leadership style of the building administrator also influences the challenges they face. Additionally, the extent to which teacher leaders (e.g., department chairs, instructional coaches, association representatives) are involved with TPEP implementation also impacts the principal’s scope of TPEP-related activities, such as organizing and leading professional development.

A high priority for district administrators in all six case study districts was to provide additional support for administrators with responsibilities for evaluating teachers. However, over a third of principals surveyed in 2014 felt that their district did not provide the staffing support that they needed to effectively evaluate teachers. Supporting principals is complicated by the nature of their work and by virtue of the fact that some school leaders struggle with delegating aspects of what they perceive to be core responsibilities to others in the building. Supports provided to building administrators included assigning additional assistant principals or instructional coaches to buildings. In addition to building level changes in staffing, all six case study districts made

staffing changes by adding or re-allocating positions at the district level to support instructional improvement initiatives. Nearly all the case study districts also described changes in hiring practices as a result of TPEP implementation.

Given the enormous effort needed to implement the revised evaluation system, some had hoped that after an initial period of adjustment, the workload issues might settle out or even decrease over time. However, in discussions with staff in the case study districts, school and district leaders reported that the workload has not decreased. Rather, as districts have been at it longer, the work has shifted and administrators find they are more intentional in prioritizing and delegating tasks. Several principals talked about using principal interns, or teachers working on administrative credentials, to support discipline and management in the school or using teacher coaches in ways that free up the principal to address other priorities.

Role of Leadership

School and district leaders in the six case study districts approached the enormous system-level changes needed to implement a revised evaluation system by promoting a conception of teaching and learning that supported continuous improvement. These leaders took the long view of implementation for comprehensive change within their local contexts, and expected that it would take a number of years for the evaluation system to be operating at an optimal level. District and school leadership actions set the tone for productive engagement with staff in the implementation efforts.

Overall, the districts tried to get ahead of the curve with messaging about the purpose of the evaluation and how their district had chosen to approach it. In some ways, the six case study districts can be located on a continuum from those with a more relaxed approach to implementation to those that attempted to do it all from the beginning. In doing so, a pattern emerged as to whether there was enough structure to support activities without putting on too much pressure that negatively impacted staff. In settings where the expectations for certain levels of performance were very high, teachers and administrators expressed more reservations about the evaluation system. In short, the notion of staff buy-in was related to both district and school leadership and how well these leaders were able to focus the initial evaluation process in ways that were manageable for staff.

Changes in teacher evaluation also illustrate the expanding role of the principalship. While this notion has been discussed in recent years, teacher evaluation brings it into sharp focus because it touches on so many aspects of instructional practice. Educators are again considering whether one person in the role of principal can be expected to carry out all the responsibilities associated with being both an instructional leader and a building manager.

Use of Electronic Tools

Technological tools have played a role in the initial implementation of TPEP. In particular, eVAL, a locally developed web-based tool has been used to support the TPEP implementation. Approximately 150 districts across Washington have tried using eVAL since its introduction, and 105 are currently utilizing it in the second year of implementation. eVAL use varies

considerably across the case study districts, with some using it to support observations, conduct self-assessments and complete scoring, and others using it primarily to store artifacts and evidence. Four of the case study districts use eVAL, while two have opted for alternative tools.

As with early versions of most new software programs, eVAL experienced initial bumps in usability which required multiple changes to the system. In the first year of implementation, principals frequently assumed eVAL data management responsibilities for their teachers, in an effort to ease the burden of this aspect of the evaluation process. In cases where it has been an effective tool and the staff use it, it is often because the principal is comfortable with technology tools and teachers are willing try the technology.

Rates of eVAL utilization by teachers differed across several variables. High school teachers were found to use the eVAL tool at higher rates than middle school or elementary school teachers. Also, those who had been teaching for less than four years were significantly more likely to use eVAL in comparison to more experienced teachers. Teachers who have tried to use the tool are fairly evenly split regarding their views of whether eVAL is relatively easy to use or whether eVAL helps them manage the evaluation process. What is clear is that teachers and administrators are becoming more familiar with the eVAL tool, but that they need more technical support and training.

Some principals credit eVAL with improving dialogue with teachers around evaluations, in terms of having access to the same pieces of evidence, being able to comment on uploaded artifacts, and being able to send questions to teachers prior to observations. Others note that work done in eVAL across the year resulted in significant time savings when they needed to generate final summative scores and reports for the teachers they were evaluating.

Views of Educators about the Impact of the Revised Evaluation System

The majority of educators participating in this study agree that the revised evaluation system has a number of benefits, but concerns about time and workload are significant. There is broad agreement that the implementation of the revised system has had a positive impact on professional conversations about what constitutes effective teaching, professional growth of teachers, and high expectations for student learning. On the other hand, nearly all teachers and principals agree that the evaluation system increases their workload, and the vast majority of building and district administrators agree that the biggest obstacle to implementation is time spent on evaluations. Nevertheless, school and district administrators are clear in insisting that they do not want any additional changes to the existing state requirements and policies because they could potentially disrupt the work underway.

Policy Implications

The following potential policy implications may help inform school and district leaders, as well as state policymakers who are considering how to support and sustain the efforts of schools and districts to productively engage staff in the revised teacher evaluation system:

Allow time for implementation of complex policy

TPEP has resulted in a massive and complicated change in school policy and practice in Washington state. The revised evaluation system requires a reconceptualization of the role of the principal and necessitates that school leaders develop deep expertise in instruction and spend substantial portions of their time observing and providing feedback to teachers. Educators in the case study districts took the implementation effort very seriously, but changes of this magnitude take years, as they touch many aspects of schooling.

Stay the course – no more policy changes

School and district administrators are clear in insisting that they do not want any additional changes to the existing state requirements and policies which could interfere with the work that is currently taking place. Educators indicate that they are becoming more comfortable with the process and developing more trust in the system as teachers work through it. Most would also like to stay the course on student growth using formative assessments.

Invest in capacity building and coherence

The case study districts found that additional administrators and restructuring of human resources were needed to conduct and support the evaluation process. Districts can also seek ways to connect the evaluation system to an instructional improvement agenda that supports a conception of teaching and learning for continuous improvement. In this way, districts can bring coherence to the many aspects of system change that are prompted by TPEP. Educators should seek to capitalize on the professional conversations with teachers about the progress of their students.

Provide continued support for integrated professional development

As the state offers opportunities for TPEP training, they should be designed in such a way as to be easily integrated with other pressing needs such as Common Core State Standards, differentiated instruction for subpopulations of students, and alternative assessments. Districts indicate a need for additional, appropriate assessments, which can support the integrity and reliability of the evaluation process, as well as training in how to use assessments to inform student growth goals. Support for and refinement of electronic tools that can support the evaluation process are also needed.

Engage support of teacher preparation institutions

The TPEP model has many elements in common with the content that is provided as part of the preparation of beginning teachers. Educators have noted that in some ways, TPEP implementation is easiest for novice teachers who have been trained in areas such as classroom assessment, use of instructional frameworks, goal setting, and evidence collection as part of their preparation program.

Introduction

This final report, prepared for the Bill and Melinda Gates Foundation and the Office of Superintendent of Public Instruction (OSPI), presents findings from the University of Washington research study of the implementation of Washington’s Teacher and Principal Evaluation Project (TPEP). This research examines the initial implementation of TPEP across Washington state districts, with a specific focus on how educators set goals and use student growth measures to demonstrate achievement of these goals. Data for this study were collected over a two-year time-period, which coincided with the initial implementation of the revised evaluation system. Through an in-depth analysis of six districts and nineteen schools, this study examines the issues Washington schools and districts face in the initial implementation of the revised evaluation system, including ways the evaluation process prompted opportunities for teachers and other professionals to engage in professional learning for the purpose of instructional improvement, and how eVAL and other electronic tools have been used in the evaluation process.

The report begins with background information on TPEP in Washington state, a brief discussion of relevant literature on teacher evaluation, and provides a description of study questions and methods. The report then provides a discussion of findings, followed by conclusions and potential policy implications.

Background on the Teacher and Principal Evaluation Project

The past decade has seen a flurry of policy objectives aimed at improving teacher quality. One of the most prominent policy goals at both federal and state levels has been to alter teacher evaluation systems. As in many states throughout the nation, Washington state adopted legislation which moved teacher evaluation systems away from a two-tier system of satisfactory/unsatisfactory to a four-tier system. During the 2010 legislative session, the passage of Senate Bill 6696 established a revised statewide system for teacher and principal evaluation. In addition to changing the state’s required teacher evaluation from a two-tier system to a four-tier system, the legislation created eight new criteria on which teachers and principals were to be evaluated. The new legislation required that districts select one of three approved instructional frameworks as the basis for a common language about quality teaching that is aligned with the state standards.¹

¹ The three instructional frameworks are: Charlotte Danielson’s Framework for Teaching, the Marzano Teacher Evaluation Model, and the Center for Educational Leadership’s 5D+ Evaluation Rubric (CEL). Descriptions of these frameworks can be found at: <http://tpep-wa.org/the-model/framework-and-rubrics/>

In 2010-11, eight districts and one regional consortium consisting of an additional eight districts developed models of evaluation systems. These pilot districts received funding from OSPI and participated as members of a steering committee advising the state on the implementation of the revised teacher evaluation system. In 2011-12 these districts piloted their evaluation models. Also in 2011-12, an additional 70 districts received Regional Implementation Grants (RIGs) to support their learning about the work of the pilot districts and to prepare for implementation of the evaluation system. An additional 138 districts received RIG support in 2012-13. In total, 208 of the state's 295 districts received some type of support to prepare for implementation. All districts were required to begin initial implementation of the revised evaluation system starting with the 2013-14 school year. At a minimum, in the first year all districts were required to implement the evaluation system for teachers who were on provisional status (those with less than three years of teaching experience) or probationary status.

The state identified two types of evaluations: comprehensive and focused. Teachers on provisional or probationary status must be evaluated on the comprehensive evaluation, meaning that the evaluation must assess all of the state's eight criteria in developing the evaluation rating. A key component of the revised evaluation system is evidence of student growth, as identified in three of the eight criterion for teachers. A subsequent bill provided additional specification about student growth measures, and mandated that student growth data must be a substantial factor in evaluating the summative performance of certificated classroom teachers. It is student *growth* in subject-matter knowledge, understandings, and skill between two points in time, not student achievement, that is relevant as a form of evidence for use in the state's teacher evaluation system. The legislation refers to the use of state tests as a possible measure of student growth, but districts are not required to use them. Current statutory language defining student growth measures is provided below:

Student growth data that is relevant to the teacher and subject matter must be a factor in the evaluation process and must be based on multiple measures that can include classroom-based, school-based, district-based, and state-based tools. Student growth data elements may include the teacher's performance as a member of a grade-level, subject matter, or other instructional team within a school when the use of this data is relevant and appropriate. Student growth data elements may also include the teacher's performance as a member of the overall instructional team of a school when use of this data is relevant and appropriate. As used in this subsection, "student growth" means the change in student achievement between two points in time. (Revised Code of Washington 28A.405.100)

The use of multiple measures of student growth in teacher evaluation adds a fundamentally new and complex feature to the teacher evaluation system. Due to its grassroots approach to developing and piloting a variety of strategies for implementing revised teacher evaluation

systems, Washington state serves as an interesting case to examine the different ways student growth measures are being used to inform teacher evaluations.

Brief Summary of Relevant Literature on Teacher Evaluation

During the past decade, there has been a dramatic increase in efforts to improve systems for teacher evaluation at federal, state, and local levels. Practitioners, policymakers, and researchers have been describing the inadequacies of traditional teacher evaluation for many years, both within the United States and across numerous nations (Isore, 2009). It has been noted that status quo evaluation systems rarely (if at all) identify teachers as unsatisfactory (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Weisberg, Sexton, Mulhern & Kelling, 2009). In addition to this very basic critique, issues such as a lack of established standards for effective teaching, minimal focus on student learning, lack of time and attention to the evaluation process, and little to no guidance for how evaluation can inform the improvement of instructional practices are cited as reasons why teacher evaluation is in need of improvement (Darling-Hammond, 2013; OECD, 2009; Toch & Rothman, 2008; Weisberg, Sexton, Mulhern, & Keeling, 2009).

While a consensus has emerged across a broad array of practitioners, policymakers and researchers that teacher evaluation is in need of fundamental change, there are widely disparate views about both the *purposes* of teacher evaluation systems and the *approaches* to designing and implementing an improved system.

Purposes of Teacher Evaluation

Simply stated, there are two fundamental purposes for teacher evaluation: accountability and improvement (Papay, 2012). Evaluation for accountability has primarily focused on using teacher evaluation to make decisions about hiring, firing, tenure or salary. In recent years, the means for conducting evaluation for accountability purposes includes determining how a teacher's performance in the classroom contributes to student learning. This implies a high-stakes system of evaluation, and one that is currently being debated, designed or implemented in numerous states (Lavigne, 2014).

In contrast, evaluation for improvement examines the extent to which both the process and the results of teacher evaluation can inform decisions about the kinds of professional learning opportunities needed to help teachers and schools engage in continuous improvement (Danielson, 2011; Goe, Biggers, & Kroft, 2012; Looney, 2011). For example, Darling-Hammond (2013) argues for a conceptualization of teacher evaluation as part of a teaching and learning system. She articulates five elements: common state standards, performance-based

assessments based on the common standards, local evaluation systems aligned to state standards, professional learning opportunities supporting quality teaching, and support structures for evaluators and others offering support.

While these two purposes of accountability and improvement represent distinctly different viewpoints, accomplishing each purpose surfaces similar questions regarding the extent to which an evaluation system is valid, reliable, and fair. Additionally, both purposes assume that evidence must be collected, although the forms of evidence used and the role that evidence plays in shaping the process of evaluation is under debate. Consequently, several approaches to the design and implementation of teacher evaluation systems have emerged.

Approaches to Design and Implementation

Recent federal policy decisions have sent strong messages to states about the ways in which teacher evaluations systems should be designed and implemented. Both the federal Race to the Top grants and the federal waivers of the No Child Left Behind Act of 2001 require that states make changes to their teacher evaluation systems as a condition for participation (Herlihy et al., 2014). Specifically, states are being asked to incorporate the results of student performance on state achievement tests to be included as a factor in the design of teacher evaluation systems.

The focus on using standardized test scores as a required part of teacher evaluation prompts important questions about how to measure the impact that teachers are having on student learning. One fundamental question concerns the extent to which changes in standardized test scores are appropriate for the purpose of gauging teacher effectiveness (Harris, 2011). The increased focus on annual testing of students, combined with both technological advancements and investments in developing longitudinal databases have provided increased capacity for tying measures of student achievement to individual schools and teachers (Papay, 2012). Thus, value-added models have become possible as a method of gauging teacher effectiveness. However, value-added methodologies are replete with numerous and thorny substantive and technical challenges which affect the development of valid and reliable estimates of effectiveness (McCaffrey, Sass, Lockwood & Mihaly, 2009; Reardon & Raudenbush, 2009). Another challenge of this aspect of teacher evaluation involves the treatment of teachers who do not teach subjects or grades in which state tests are available (Goe & Holheide, 2011). Nevertheless, a number of states have moved forward with the design and implementation of teacher evaluation systems that incorporate measures of student academic performance into the evaluation process (Steele, Hamilton, & Stecher, 2010). This does not mean that student test scores are the sole measure in teacher evaluation systems. On the contrary, multiple measures of teacher effectiveness are being incorporated in the design of new systems. Steele, Hamilton & Stecher (2010) describe two reasons for relying on multiple measures: (1) they improve the completeness

and accuracy of judgments about teacher effectiveness, and (2) they address the issue of non-tested grades and subjects.

Implementing teacher evaluation systems that use multiple measures of student performance implies that variation will exist in the specific measures to be used. This holds irrespective of whether the purpose is primarily for accountability or for improvement. Options being considered as additional measures of teacher effectiveness include classroom-based assessments, benchmark assessments, portfolios, student surveys, and observational assessments, including peer review (Bill and Melinda Gates Foundation, 2013; Darling-Hammond, 2013; Goe, Biggers, & Croft, 2012; Humphrey, Koppich, Bland & Bosetti, 2011; Steele, Hamilton, & Stecher, 2010). States and districts that are incorporating multiple measures that include student performance in the design of their evaluation systems are in the process of examining ways to ensure validity and reliability of scoring systems. However, a set of strategies for addressing the challenges of reliability, validity, and training of evaluators has not yet emerged (Herlihy et al., 2014).

Research that examines how states and local districts are responding to the dramatic changes in teacher evaluation systems for purposes of both accountability and improvement can serve to inform both design and implementation efforts that are underway.

Research Questions and Methods

This research study is a mixed methods examination of the issues districts and schools face in implementing the state's revised evaluation system with regard to goal setting around student growth, and the identification, selection and use of appropriate data to monitor, adjust and evaluate the achievement of these goals. We also explore the ways in which an emphasis on the student growth rubrics in the evaluation system may help support professional learning. The overarching research questions include the following:

1. What kinds of student growth goals are selected by teachers in the evaluation process and what informs those decisions?
2. What forms of evidence are collected to document student growth and in what ways is the focus on student growth shaping improvement strategies?
3. In what ways has the evaluation process prompted opportunities for teachers and school leaders to engage in professional learning around goals and measures of student growth for the purpose of instructional improvement?

4. What actions have educational leaders taken to support authentic engagement in the evaluation process? What challenges do educators face in implementing the revised evaluation system?

5. In what ways is eVAL (and/or other electronic tools) used in the evaluation process?

Methods

To address these questions, the study employed a concurrent mixed-methods research design (Creswell & Plano Clark, 2011) with both qualitative and quantitative data collection in a strategic sample of six districts, including case studies of 19 schools within these districts. The districts selected vary by size, regional location, type of instructional framework used in the evaluation process, implementation timeframe, proportion of novice teachers, and the demographic characteristics of the students served. The districts were also identified as early adopters of TPEP and constructively engaged in implementation activities through pilot and RIG participation. The six districts are Cashmere, Evergreen, Kelso, Kennewick, Mount Vernon, and Renton. Demographic characteristics of the study site districts are displayed in Table 1.

Table 1: Characteristics of Study Site Districts				
District	Enrollment*	Poverty Rate* FRPL	Framework	%Novice (<4yrs) 2013
Cashmere	1,491	50.9%	CEL	9.6
Evergreen	26,486	47.2%	Danielson	15.3
Kelso	5,018	58.5%	CEL	14.7
Kennewick	17,024	53.2%	Danielson	17.1
Mt. Vernon	6,402	67.3%	Marzano	22.3
Renton	14,902	54.6%	Danielson	25.6

**Based on May 2013 Student Count from OSPI's Washington State Report Card*

The primary qualitative strategy involved semi-structured interviews with teachers, principals, and district staff that were conducted during the 2013-14 and 2014-15 school years, and collection of relevant district and school documents. The main quantitative work utilized online surveys completed by teachers and other certificated instructional staff, as well as school and district administrators who were involved with teacher evaluations in these districts. Surveys were administered at two points during the study period: Late Fall/Winter of 2013 and Fall of 2014.

Interview Data

The first round of site visits occurred in the Summer and Fall of 2013 and focused on generating a detailed description of the early implementation efforts of the evaluation system in these districts. These interviews were used to inform the development of the first online survey which was deployed in late Fall 2013, in five of the six districts. The second round of site visits occurred in Spring 2014 in all six districts. A third round of interviews (as well as site visits in four of the six districts) were conducted in Fall 2014. District and school administrators associated with TPEP were typically interviewed two or three times, and teachers in the case study schools were interviewed once during this phase of the research. Overall, 26 district administrators and staff, 33 school administrators, and 123 teachers and instructional staff were interviewed, resulting in a total of 214 interviews over the two-year period. A follow-up survey was administered in Fall 2014, to examine changes in educators' perceptions of the TPEP implementation over the two year time period. In addition, a variety of archival sources were collected to offer both qualitative and quantitative information pertinent to the research questions, especially with regard to the organization and focus of professional development, demographic information about students and staff, and the leadership efforts aimed at addressing goal setting and evidence of student learning.

Survey Data

Survey data from school and district staff in the case study districts offer a broad source of information concerning implementation efforts and challenges. These include factors such as the nature of goal setting around student growth and corresponding forms of evidence, the rationale for these decisions, professional development activities, use of technological tools and potential alignment with other school or district improvement initiatives. Surveys were constructed in consultation with the districts. The surveys have a common set of items for all participating districts, but also items specific to the individual district context. This design allowed for the collection of information pertinent to the study purposes and also for formative feedback to the individual districts.

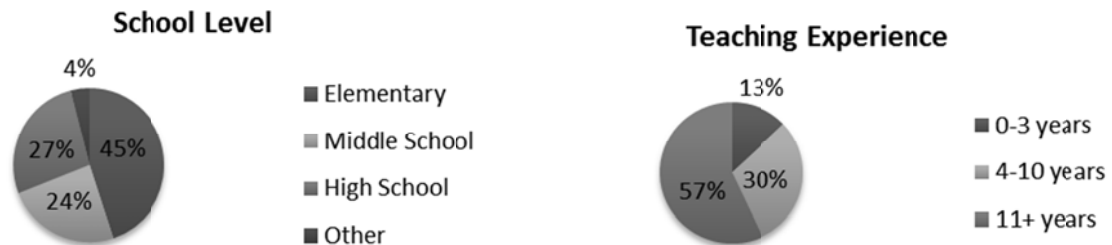
Surveys were sent to all certificated teachers and administrators in each district. Survey participants were asked to identify the role that best describes their current position. The survey was branched based on the identified role and each group received a unique set of survey items. Items were also branched within the survey for each group, depending on how participants answered particular sets of questions. Across the five districts, a total of 1,779 individuals responded to the initial Fall 2013 survey. The vast majority of respondents identified themselves as teachers (86%). An additional five percent identified as school-based or district-based instructional coaches, and the remaining nine percent identified as a school principal, assistant

principal, or district level administrator. The overall response rate was 43 percent. Response rates by district ranged from a high of 90 percent to a low of 33 percent.

The follow up survey in Fall 2014 was sent to certificated instructional staff and administrators in all six case study districts, with 2,024 individuals responding. As in the prior year, the vast majority of respondents identified themselves as teachers (86%). The overall response rate was 44 percent. Response rates by district ranged from a high of 78 percent to a low of 31 percent.

Survey participants who identified as teachers (including those who identified as instructional coaches or specialists) represented a range of teaching experience and grade levels, as shown in Figure 1. Of the teachers who responded in 2013, an average of 61 percent reported that they participated in an evaluation last year as part of their district’s pilot efforts. However, responses to this item varied by district. By 2014, eighty percent of all teachers were participating on either a comprehensive or focused evaluation plan. With respect to other involvement in TPEP, a small portion of responding teachers and instructional coaches (13 percent) indicated they had served in some formal role in the TPEP process this year or in past years. These roles included serving as a member of a school or district committee, a workshop leader or presenter, or a teachers’ association representative for TPEP in their school or district.

Figure 1: Teacher Survey Participants by School Level and Teaching Experience



In addition to the survey data, we collected de-identified information about teachers’ goal setting on the student growth criterion, and the forms of evidence used to measure that growth. For some districts, the 2013-14 year was the first time teachers have worked on goal setting and evidence of student learning in this process. Since there is considerable variation in the use of electronic tools to facilitate the goal setting process, the nature of data collection depended on the practices, tools and documents used by individual districts. We also received data from the eVAL developers that provided statistics regarding eVAL use by districts and schools. These were used to triangulate reported use by case study districts.

Examining the Early Implementation of TPEP: Findings

In this discussion of the study findings, we examine how the six districts approached their implementation work. Overall, these findings suggest that the early implementation of the revised evaluation system has supported or created some conditions for the improvement of teaching and learning, but also resulted in additional challenges for local school systems. Among the six districts, there is considerable variation in how they have approached the implementation, as well as the professional development needed to support staff through this process. One of the most challenging aspects for educators has been the introduction of student growth goals and the collection of evidence aligned to the goals.

Introducing a change as massive as a revised state-wide evaluation system necessarily involves educators across multiple levels of the system. Restructuring human resources to accommodate the increased workload and costs, particularly to assist principals and others in evaluative roles, has resulted in shifts in staffing and some unanticipated consequences. Technological tools in the roll out of the evaluation have played a role in the initial implementation, though perhaps not as large as initially envisioned. In short, the revised evaluation system impacts and touches numerous aspects of the school system.

Variation and Innovation in Early Implementation (Years One and Two: 2013-14 and 2014-15)

Selection of the instructional framework

Statewide, districts are roughly split in thirds when it comes to their selected framework, with approximately 36 percent of districts adopting Danielson's Framework for Teaching, 34 percent selecting the CEL5D+ Teacher Evaluation Rubric, and 31 percent using the Marzano Teacher Evaluation Model. Many of the state's larger districts selected the Danielson Framework, while a sizeable portion of smaller districts adopted the Marzano model. Because of this variation by district size, a majority of the state's students are in districts using Danielson's Framework (53 percent) and proportionately fewer are in Marzano districts (14 percent). Among the six case study districts, three selected the Danielson Framework (Evergreen, Kennewick and Renton), two selected CEL 5D+ (Kelso and Cashmere) and one selected Marzano (Mount Vernon). According to the districts, the selection of the framework was influenced by prior experience or familiarity with a particular framework, regional supports available, or interest in working with local educators on these issues.

All the case study districts were early implementers of the revised teacher evaluation system. All sought grant funding from the state to pilot aspects of TPEP (as part of an initial set of pilot districts or Regional Implementation Grants – RIG I), and all were collaborative in seeking to

foster buy-in and ownership from a large group of diverse stakeholders. In most cases, the superintendent took direct responsibility to lead the effort in a high profile manner. In a few districts, the assistant superintendent for teaching and learning or another district-level leader carried the charge. These leaders were careful to include school and district administrators, teachers and teacher leaders, and teacher association representatives in discussions and decision-making regarding the adoption of instructional frameworks, professional development, and assessment issues. Both pilot and RIG I districts were required to have association leaders as members of their team.

In their selection of the Danielson Framework, Evergreen Public Schools formed a TPEP committee which included the perspectives of diverse stakeholders such as district and building administrators, teachers, and association representatives. Ten years prior, Evergreen had introduced the Danielson Framework in their instructional improvement efforts, which influenced their decision to remain with the framework. However, in the district's early work with the Danielson Framework, they chose to focus on just four of the 22 criteria (engagement, assessment, high expectations and professional collaboration). Once they became a RIG I district, they adopted the full set of indicators within the framework. While TPEP committee members acknowledge considering the other two framework options, they decided to stay with the Danielson Framework because many teachers and principals were already familiar with the rubrics, and because switching frameworks didn't seem to offer any real advantage. Despite familiarity with the framework, Evergreen leaders recognized that additional training was still required to help educators gain a more in-depth understanding of the framework. Training was also necessary for staff who had not received the original training and those new to the district.

For more than a decade, the Kennewick School District had been developing their own instructional improvement model, which focused on *purpose, engagement, rigor* and *results*. Using these four components as a guide, the district provided principals with multiple opportunities throughout the year to analyze video-taped lessons of teachers and to discuss instruction. A district administrator explained that the work they had done in developing their own instructional framework was one of the reasons they were interested in participating in the original pilot work around TPEP. In the first year of the pilot, they took the model they had developed and tried to fuse it with the Danielson Framework because it had a bit more detail. But when the state decided that all districts must select one of the three frameworks (Danielson, CEL 5D+, or Marzano), Kennewick abandoned their locally-developed model in favor of the Danielson Framework because it was closest to what they had developed.

The Cashmere School District organized their TPEP implementation around a large team, which included teacher leaders, principals and administrative staff, and a school board member. As a small district, educators reported that they all had an active part in the framework decision. Prior to receiving their RIG I funding, the district had explored the CEL 5D+ instructional framework

which, according to the superintendent, resonated with them for multiple reasons. First, they felt their prior professional development around “Powerful Teaching and Learning” with the BERC Group was aligned well with CEL 5D+ and what they were already doing. Second, the district liked the five dimensions and thought that by focusing on one or two dimensions at a time, they could introduce the framework gradually. Third, they liked working with an in-state non-profit organization, and thought an in-state research-based model could help support the evaluation work. The leadership team visited other districts across the state and listened to presentations on the three frameworks before making a final recommendation to the school board.

The Mount Vernon School District selected the Marzano Teacher Evaluation Model due in large part to the district’s three year participation in the NW Regional Summit Grant. The district decided that continuity with the framework was important both for administrators and teachers who had received exposure to Marzano-based research on improving student learning. Building administrators also had completed three years of classroom walkthroughs using the Marzano framework. Mount Vernon organized professional development activities specifically around the Marzano framework. Principals and assistant principals attended training in Wenatchee so they could observe a TPEP rollout using the Marzano framework, and teachers had the opportunity to attend summer training around rater agreement on scoring the Marzano rubric at Western Washington University.

In the Kelso School District, a steering committee comprised of thirteen administrators and teachers worked with the district in the early implementation of TPEP. This committee examined the three different frameworks, ultimately choosing the CEL 5D+. According to a principal who sat on the committee, some district level educators had had experience with CEL 5D+, which “kind of helped lead our decision.” This principal, who had previous experience with the Danielson framework, did not feel that the CEL 5D+ offered a measurably different experience with regard to evaluating his teachers. He explained, “I don’t know if it really matters in the end. It’s nice having a rubric and to be able to tell people what you’re looking for.”

As a RIG I district, the Renton School District formed a very large TPEP committee with school and district representatives from across the district and within each school, including association representatives. These 100 or so representatives met to discuss the three instructional frameworks and then share this information with their colleagues. The Danielson framework was selected because it was most closely aligned to the district’s ongoing instructional work. Renton had used the Danielson framework in some professional training prior to TPEP, and they also had senior staff with experience as trainers of the model.

Focus on instructional improvement and growth rather than compliance

Among all six case study districts, educational leaders directly connected the evaluation system to an instructional improvement agenda. They adopted a learning stance towards teacher

evaluation, emphasizing the potential benefits rather than viewing the revised system as yet another state mandate requiring compliance. One superintendent described the work of the district's TPEP committee in the following way: "As a group we said, 'OK, this is the way the state is going.' Our agenda, I think collectively has been 'how do we do this, and do it well...'. Our whole focus has been 'how do we help our teachers get better?'" In all cases, the message was one of using the evaluation as an opportunity for instructional improvement and conversations about effective teaching and student learning. Another superintendent explained, "There should be urgency around improving teaching and learning. The urgency shouldn't come from a state law. It should come from within the system."

Teachers also noted the ways in which the required change to a revised evaluation system represents a focus on growth and an improvement over the past. A teacher leader at a middle school talked about changes to the evaluation system that she saw as positive: "Most people would agree that the old evaluation system was not helpful for growth. Even though this one – at first people get a little anxious about the new system or anything new – I think everyone would agree that it's a growth system and I think people do feel supported. And the message has been: This is not a gotcha! This is about, how do we work together to further student learning?" The continual focus on the improvement of instructional practice is why one superintendent thinks his teachers' association isn't worried about negotiating the contract language around TPEP. Teaching involves a very complex set of tasks, and this superintendent talked about how all educators can improve, and that no one is 'distinguished' all the time. He shared: "We've crossed the bridge saying it is okay to be 'basic' in some areas – you're not going to be distinguished in every area." Having a growth mindset has enabled them to constructively engage with staff on these issues.

These findings suggest that the use of frameworks in guiding instructional improvement efforts has provided a rich medium for conversations about instruction. Survey data from both 2013 and 2014 confirm that a majority of teachers and building administrators find the instructional frameworks adopted by their districts to be a good model for effective teaching (85 percent of teachers, and 99 percent of building administrators, agree either somewhat or strongly), and provide a common language to discuss effective instructional practices (87 percent of teachers and 98 percent of building administrators either somewhat or strongly agree).

While these conversations could create opportunities for deep and lasting impact on instruction, a challenge among all the districts was the substantial investment of time this required. A district administrator explained,

"If we do this right, we can impact instruction with TPEP. If we make this a compliance issue, you can get it down and you can manage it, and everybody goes on. But if we do it right it, we can change instruction, improve instruction across the school district. But

that takes time. And it is the conversations. So it's a double-edged sword. People say that the most powerful thing is – we have these great conversations with teachers because we have this rubric as the center of the conversation. But the most challenging thing is the pre and post conferences take 45 minutes to an hour a piece because we have this rich body of information in front of us.”

Educators recognize that the teacher evaluation policy has the potential to fundamentally change how teaching and learning happen, but also that it requires framing the conversation in ways that productively engage staff in support of authentic, long-term professional growth.

Early work with teacher leaders

Whether as a pilot or RIG district, the goal was to learn more about the evaluation system, select an instructional framework, and pilot the evaluation with a portion of teachers ahead of the formal implementation. All of the case study districts were strategic in how they approached introducing TPEP to their staff. A common theme among the districts was to invite a few well-respected and experienced teachers to participate in the early evaluations. A superintendent explained, “If we can get some good veteran teachers involved in the comprehensive evaluation and they see how that goes ...they can be advocates; they can support their colleagues.” One superintendent described decision making around TPEP in terms of a team of teacher leaders who piloted the system. However their primary focus was on defining and recognizing good instruction, not necessarily on the specific implementation issues involved with TPEP.

As one of the original pilot districts, Kennewick took a lead group of teachers and put them on a comprehensive evaluation four years ago. A building administrator explained, “Once you get a lead group comfortable, they tend to spread the word pretty well and then you can start adding them [teachers] on... It was a lot easier than I thought. Now we don't have to battle over it, and it ripples out into the staff.” As a RIG I district, Evergreen piloted the system with approximately 25 percent of its teachers on a comprehensive evaluation, including all provisional teachers, and a number of veteran teachers from across the experience continuum. By piloting the revised evaluation system with well-respected veteran teachers, these districts not only developed advocates for the revised system, which helped to elicit buy-in among teachers across these districts, but they also created a cadre of teachers who could then support their colleagues as they moved through their own evaluations.

Strategic decisions for transitioning to the revised evaluation system

For the initial state rollout in 2013-14, the case study districts adopted different implementation approaches for transitioning their staff on to the revised evaluation system. Four of the six districts decided that virtually all teachers would be evaluated on either a comprehensive or a focused evaluation in the first year of statewide implementation. The other two districts chose to stagger staff participation over three years. Educators in these districts mentioned two main

reasons for employing a staggered implementation. First, they described considerations such as having large numbers of teachers on provisional contracts, which would mean placing them on the more labor intensive comprehensive plan. Second, they cited negotiations with their teachers' association as a factor in the decision.

In the first year of statewide implementation, Cashmere moved nearly all teachers in the district to the revised evaluation system. Most of the teachers were on a focused plan (approximately 75 percent) with the remainder on a comprehensive plan, and all principals and assistant principals were involved in evaluating teachers in that first year. Kelso also decided to implement TPEP across the board in the first year so that every educator in the district would be evaluated. According to a member of their TPEP steering committee, this decision was positive for teachers, but a burden for administrators: "From the teachers' end of it, I think it was a good idea. But boy do I feel sorry for administrators."

In one of the districts, a gradual transition plan to TPEP was part of a negotiated memorandum of understanding with the teachers' association. Half of the teachers in the district were to be evaluated in the first year of formal implementation, increasing to 75 percent in the second year, and finally all teachers by 2015-16. In retrospect, a superintendent in one of the districts where a phased-in approach was adopted indicated that their principals wished they had moved all teachers to the revised evaluation from the beginning. Even though 75 percent of teachers were on the revised system by the second year, staff were not all on the same page about the implementation process, which created additional stress and anxiety. Educators in the other district utilizing a staggered approach expressed similar misgivings. A member of the TPEP committee explained that at the time, a phased-in approach seemed like a good idea. But, she went on to explain that attempting to provide professional development when only a portion of the staff were actively engaged in the process resulted in uneven communication and support, and some staff feeling left out of the process.

District and school approaches to introducing the framework and professional learning

The case study districts all used familiar delivery mechanisms to provide support to teachers and principals through early intensive phases of professional development. But several districts did not require that teachers be evaluated on all the framework components in their evaluation plan in the beginning. District leaders explained that this decision was due in part to a lack of clarity and guidance from the state on certain aspects of the evaluation, but also because they believed their staff needed more time to become familiar with the framework components and the overall evaluation process.

One district decided to "chunk up" staff learning of the instructional framework, making a conscious decision not to try and introduce it all at once. Their strategy was to focus on specific dimensions of the framework throughout the rollout period. As a RIG I district, the staff focused

on the dimensions of *purpose, classroom environment, and culture*. In the next year, staff concentrated on *student engagement*, and in the following year, they planned to add *assessment*. The superintendent saw this as a way to layer and build upon teachers' understanding and reduce stress and anxiety. By adding different dimensions every year, they were able to continue to monitor and further develop those introduced earlier. The director of teaching and learning noted, "The alignment of those three areas is aligned pretty well to the 'State Eight.' So we can still evaluate teachers pretty thoroughly, even though we haven't covered the entire framework...and we thought it was better to go deeper into each dimension with all of our staff."

The superintendent was very clear that they wouldn't evaluate staff on dimensions for which they hadn't yet provided professional development support. A building principal clarified, "We are really going to focus on holding people accountable to what we supported and not so much about what we haven't supported. For example, we have not delved into assessment so much. So, we're not going to be looking at that framework [element], and expecting people to be well-versed in those expectations until we've actually supported their growth in that." With regard to assessments and the Common Core, the superintendent further explained, "We purposefully didn't choose assessment [this year] because we don't know what the standards are yet [for Common Core]. How do you assess something when you don't know what the targets are? We're just taking it a chunk at a time." Due to an incremental roll out, there had been considerable time devoted to learning certain framework elements, but less attention to goal setting for student growth in this district.

Similar to the notion of districts incrementally introducing the framework and targeting the evaluation around those dimensions, this study also found that some districts and schools chose to selectively focus professional learning on specific aspects of the framework, even while they evaluated teachers across all the dimensions. In these cases, educators focused on the framework dimensions that were most salient to their day-to-day professional realities. For example, an elementary principal using the Danielson Framework chose to focus solely on Domain 2 components (the classroom environment) in the first couple months of the school year. While she had originally planned to focus on four components per month so she could cover all 22 components during the first half of the year, she adapted her plan so that she could focus on the domains that were in alignment with her most pressing priorities. Having five brand new teachers in her building, she naturally found she was focused on Domain 2 in supporting their classroom management procedures. She explained that she intended to focus on Domain 3 next, and work her way through the remaining domains during the academic year.

Building-level strategies for introducing staff to the revised evaluation system

Across the case study districts, individual schools also found innovative ways to introduce their staff to TPEP. School leaders were able to leverage existing staff capacity to support TPEP implementation in decisions such as who should initially be on a comprehensive or focused

evaluation. What follows is a brief description of strategies that building leaders used to support the implementation.

Selecting teachers by grade level. In one middle school, the principal chose to implement TPEP with his staff by intentionally choosing grade levels with very strong and tight knit Professional Learning Communities (PLCs). Because these PLCs already functioned well collaboratively, the teachers could create their student growth goals together as well, and serve as a model for future staff evaluations in the building. In another district, an elementary principal decided to put her teachers on the comprehensive plan by grade level groups (e.g., grades 2-3 in the first year, K-1 in the second year, and 4-5 in the third year). The principal reasoned that by having all or most teachers together on a comprehensive plan by grade group, she could easily meet with them as a group to support professional development on TPEP. These grade level teams of teachers naturally shared things in common, such as lesson planning and curriculum, and they also worked on student growth goals together. While this plan has worked well in the first few years, as experienced teachers move to a focused plan, and new teachers are added on a comprehensive plan at different grades, the system is starting to become more heterogeneous going forward.

Selecting strong veteran staff to be on the comprehensive plan. As we found in the piloting phase of TPEP, building leaders also chose well-respected and experienced teachers for the comprehensive evaluation early in the statewide implementation. An elementary principal deliberately chose his leadership team, which he considered his highest performing teachers, to be on the comprehensive evaluation from the beginning. He intentionally avoided choosing naysayers or those with doubts. His rationale was, “I felt like these powerful teachers, if they could stand up and say it’s good, it worked; the strong union people would say, ‘that’s one of my colleagues, I’ll do that, I can do that.’”

Use of veteran staff for mentoring/teacher leadership of junior staff. As previously noted, some districts piloted the evaluation with well-respected, experienced teachers, which according to these districts, helped to define and recognize good instruction. In doing so, including strong veteran staff early in the implementation process also had the benefit of enabling those who served in formal or informal mentoring roles to be available to support their junior colleagues. According to one district’s TPEP steering committee, a big success has been their focus on using teacher leaders: “I think that our teacher leader model has been fabulous.” In this district, the use of teacher leaders was an intentional strategy used to support other teachers and also elicit buy-in among all teachers.

Understanding the teacher workforce within a building. By law, provisional teachers are on a comprehensive plan for at least three years. However, schools with higher turnover, and often more novice teachers, may be differentially impacted by the portion of their teaching staff on a comprehensive evaluation in any given year (workload issues will be discussed in a subsequent

section of this report). When building administrators have a bit more latitude, they often make a different set of decisions. For example, in one middle school, where the teaching staff are primarily veteran teachers, the principal chose to select teachers for the focused or comprehensive evaluation based upon who he thought needed the extra support, and in some cases, pressure to improve. He also reported not putting too much focus upon teachers who were about to retire, explaining, “where should I put my energy...probably shouldn’t be in somebody that’s either going this year or next year.” The revised evaluation system provided an impetus, in a few cases, for retirement.

Videotaping teacher lessons to support instructional improvement. In a middle school that is sharply-focused on instructional improvement, the principal encouraged staff to be videotaped teaching so that he could provide individualized guided feedback on components of the Danielson Framework applied to their own practice. As uncomfortable as this can be, the principal explained that he didn’t require teachers to be videotaped, and that all but one teacher agreed: “It’s so powerful to watch yourself, probably one of the best ways to get better is to view yourself and then use that framework in a way that you have to understand it well in order to apply it to yourself.... It has a lot of potential for helping people grow.”

Shifting orientation from use of frameworks to State 8 criteria

As statewide implementation of the revised evaluation system has proceeded, in some cases, the role of the comprehensive instructional framework and the requirements of the eight state criteria (State 8) by which teachers are scored and evaluated have not always been clear to educators. The frameworks are intended to be the mechanism for the analysis of instruction, and the State 8 serve to organize and quantify the evidence gathered in that analysis for purposes of scoring. Most of the case study districts spent considerable time introducing their selected framework and then trying to do the “crosswalk” with the State 8. A teacher explained that in her district it is expected that teachers will use the Danielson framework in their PLC work. However, because teachers must now cross reference the framework with the state criteria, the process has become more complicated, and the state criteria have been given a larger focus. In many ways this emphasis is to be expected since the bottom line in the evaluation is how the rubric will be scored. However, some worry that shifting focus onto the scoring rubric will diminish the role of the instructional framework in the evaluation process.

In the second year of statewide implementation, interviews with building administrators revealed that some are using the State 8 with their staff, in some ways in lieu of the instructional framework. One district superintendent felt that they hadn’t spent enough collective time introducing the instructional framework and the connections to the State 8 criteria. As a result, some staff had chosen to focus on the State 8 rubric. The superintendent reflected, “If I had it to do over again, I would do more of an intentional crosswalk and roll it out all on the same page with the framework and the State 8.” A district administrator who led the TPEP implementation

in his district felt the focus on the State 8 rather than the instructional framework was a mistake. He explained that the State 8 aren't organized in a way that can be taught in support of instructional improvement efforts. He described the eight criteria as "random acts of teaching and no more." In this early stage of implementation, the extent to which the State 8 will be given a larger focus in relation to the instructional frameworks remains unclear.

Goal Setting and Collection of Evidence

During the first two years of statewide implementation of TPEP, goal setting and the collection of evidence for student growth are proving to be among the most challenging aspects of the revised evaluation system. Findings suggest that teachers struggle with setting goals for student growth and linking these growth goals to initiatives aimed at improvement, while principals find it difficult to allocate the requisite time to adequately help teachers develop measures to assess student learning. In this section we discuss some of the efforts and main challenges surrounding goal setting for student growth, the collection of evidence, and the need for appropriate formative assessments to support this work.

Goal setting for student growth

In the initial implementation of Washington's teacher evaluation system, the state has placed an emphasis on the process of identifying and setting goals for student growth, and allowed for varying levels of choice in making these determinations. There are five components designated for student growth embedded within three of the state's eight criteria.² Teachers on the focused evaluation plan are evaluated on one of the eight state criteria, and also must include the student growth component from one of the three criteria with this focus.³ Teachers on a comprehensive plan are evaluated on all eight state criteria. As established under the criteria, goal setting for student growth can address a subgroup of students, a whole class, or can be done as a collaborative effort in working on shared goals within grade levels, Professional Learning Communities (PLCs), subject matter teams or school-wide improvement teams.

Given that the state statutes allow for local discretion regarding what counts as evidence of student growth for purposes of the evaluation, the case study districts adopted a learning stance which created space for teachers to consider alternatives. One superintendent explained the shift in thinking as they focused on the collection of evidence:

"Our goal was really to learn how to gather evidence in a different way than we had been gathering evidence before. And what does the evidence mean? And then I think the

² These components in the criteria are 3.1, 3.2, 6.1, 6.2, and 8.1.

³ If the teacher chooses criterion 3, 6, or 8, they must complete the corresponding student growth components. If the teacher chooses criterion 1, 2, 4, 5 or 7, they must also complete the student growth component in criterion 3 or 6.

real shift we tried to make last year.... was really moving away from the concept that it is the administrators' job to evaluate the teachers and [instead] it's a partnership where the teacher brings evidence to the table and the administrator gathers evidence in other ways. And the evidence isn't what I saw in a thirty minute classroom visit. The evidence is the number of things that come together over time to have a complete look at the performance."

Early in the implementation, some districts hesitated in designing professional development opportunities around goal setting in hopes that the state would provide greater clarification about their expectations. While somewhat greater specificity has been provided, districts still report that the goal setting component has been one of the most challenging aspects of the TPEP implementation so far. One superintendent reported that moving toward goal setting may be "the key shift in whether we are successful or not in the long run." In fact, he believes that goal setting is the element in the evaluation system that will help move instructional improvement forward, emphasizing that if educators "don't set that goal and we can't measure those goals, we are back to where we've been."

One district described initially trying to develop a goal setting process with the teachers by asking them to write a SMART goal on a specific subject area. The district administrator explained, "The rules haven't come out specifically around growth yet. So, we just asked teachers to write a couple of goals. And they needed to be SMART goals related to TPEP. And they were awful. Because we expected people to do something [that] they weren't trained to do. Teachers didn't know what SMART goals were. And so that kind of fell on its face a little bit." Over the course of several years, and in collaboration with their teachers' association, the district developed a teacher goal setting form that has all the legally required components of a student growth goal. Since then, district administrators indicate that teachers' goals for student growth have been much better.

Some districts encouraged teachers on a comprehensive plan to "nest their goals" for greater coherence. A district administrator explained how he believes that the "power" for student growth comes in working collaboratively around goal setting and, where possible, to nest their growth goals. He explained:

"Let's take the example of a third grade teacher. [For] a third grade teacher, her goal, let's say, is for all the kids to know their multiplication facts. That could be one of their goals. And so, in student growth goal under criterion 3, the target is the group of kids performing below grade level. So, what we'd want to see in their growth goal, it could be around math facts, but how they are working with kids below grade level.... Then in [criterion] 6, it's really a growth goal for the whole class. But, in the way we're looking at this, it can still be around math facts. Because now it's just, with the first goal, it's how

you're working with kids who struggle in that area, the second goal is how you are teaching it to your whole class. And then [criterion] number 8, if we could really get fancy, and here is how I'm working with my team to do the same academic goal. So, it's possible to tie it all together and make a really powerful growth goal for a group of teachers, is what we're thinking."

As part of the 2013 survey, teachers were asked to provide samples of goals, and over three quarters of teachers included sample goals in their responses. An analysis of these sample goals revealed substantial variation across teachers. Some teachers provided very specific goals that were fairly narrowly defined (e.g., "increase fluency from 25 words per minute, to 30 words per minute by 3-1-14 as measured by running records and fluency probes"), while others shared very general goals (e.g., "better conceptual understanding through differentiated instruction"). Teachers also varied in the extent to which student behaviors were a focus of their goals, such as increasing student engagement or student ownership of their learning. Based on the survey data, it was evident that in the first year of statewide implementation, many teachers were still learning how to formulate meaningful goals for student growth for the purpose of the evaluation.

Resources and supports for goal setting

In the 2013 survey, teachers were asked to rate the usefulness of several sources of information in informing their goal setting for student growth. A majority of teachers (54 percent) identified classroom assessments that they developed to be very useful in informing their goal setting and an additional 39 percent found them to be somewhat useful. Nearly half of teachers (47 percent) identified conversations with other teachers about the progress of their students as very useful, and an additional 41 percent found this to be somewhat useful. A significantly smaller proportion of teachers identified state tests or other standardized tests to be either very or somewhat useful in informing goal setting. Only 6 percent reported that state tests were very useful and 13 percent found other standardized tests to be very useful. A few differences were noted when comparing teachers' responses by school level. Seventy-eight percent of elementary teachers found school or district created assessments to be very or somewhat useful in informing goal setting, compared to 49 percent of middle school teachers and 39 percent of high school teachers. A higher percentage of elementary teachers (53 percent) reported that other standardized tests such as MAP and DIBELS were very or somewhat useful, compared to 37 percent of middle school teachers and 22 percent of high school teachers. Table 2 provides additional data about the sources of information for informing goal setting.

Table 2: 2013 Survey: Teachers' Ratings of the Usefulness of Various Sources in Informing their Goal Setting for Student Growth (n=1429)

	Very useful	Somewhat useful	Not useful	Did not use/Not applicable
Conversations with other teachers about the progress of my students	47%	41%	8%	4%
Classroom assessments I developed	54%	39%	3%	3%
School or district created assessments	18%	42%	22%	17%
State benchmark data (e.g., math or reading)	15%	41%	22%	21%
State standardized Tests (MSP or HSPE)	6%	27%	28%	37%
Other standardized tests (e.g., MAP, DIBELS)	13%	27%	19%	39%

The challenge of figuring out how to practically work out goal setting with staff does not diminish the fact that most educators involved in TPEP believe that paying attention to student growth provides some benefit to the teacher evaluation process. In the 2014 survey, three-quarters (76 percent) of teachers agree either somewhat or strongly that examining student growth is a useful part of evaluation. Teachers also indicate that they have been given some supports to help with goal setting. Approximately four-fifths of teachers either strongly or somewhat agree that they have been given time to work on student growth goals (80 percent) and are able to link student growth goals to school or district improvement initiatives (76 percent). Table 3 describes these perspectives in greater detail.

Table 3: 2014 Survey: Teacher Views on Specific Supports and Evaluation Issues (n=1732)

	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
Teachers in my building have been given time to work on student growth goals	38%	42%	15%	5%
Teachers in my building link student growth goals to school or district improvement initiatives	28%	48%	19%	5%
Examining student growth is a useful part of teacher evaluation	32%	44%	15%	8%

Building administrators identified their work with teachers on goal setting for student growth as a support to teachers. Eighty-five percent of principals and assistant principals felt that teachers had received some or considerable training on goal setting from the school or district. A majority of school administrators also indicated that their teachers had received some or considerable support in working with their principal or assistant principal on goal setting for student growth (92 percent), and that their teachers had at least some support to work in their PLC or subject matter or department team writing goals for student growth (87 percent).

How and what teachers are collecting as evidence of student growth

Districts have struggled to determine the best measures for determining student growth. The collection of evidence, and the role teachers play in gathering evidence and demonstrating student growth have been among the biggest concerns related to TPEP implementation. For most districts, even those with prior experience using instructional frameworks, there had not been a prior focus on using student evidence. According to one superintendent, “that’s been a bigger issue for us than anything.” The association representative in the same district similarly explained, “I think that there is a big issue with teachers feeling that ‘what are artifacts?’ And ‘what constitutes growth,’ ‘showing growth?’” An elementary principal in a Title I school explained how he approached the challenge:

“We just sat down with our CEL5 leaders and we said, ‘What is assessment in our school? What does it look like? What is it? How do we monitor it? How do we track [student growth]?’ We generated what we felt like was a comprehensive list for the elementary school. From that, they treated it almost like a menu. Our agreement was, ‘what are we going to do with the teacher who says, I’m going to use my own pre and post-test,’ and ‘I’m going to use another classroom-based assessment.’ ... The group said, ‘Let’s make a recommendation that you have to use a district-approved assessment.’ So they have to use some larger test. That’s one point of assessment. And no one really balks at it because that is how we qualify kids for Title I, and do interventions. So we know they know the data, we know they know how to use it.”

But the principal is quick to add that he doesn’t require everyone to use the same assessments. Instead he looks at what each teacher is doing and has an individual conversation with them about what is most appropriate in their situation.

Several survey items inquired about aspects of how teachers are responding to the need to collect evidence of student growth for purposes of their evaluation. The vast majority of teachers plan to use classroom-based assessments, pre/post unit tests, or assessments developed by their department, grade level or PLC. School and district administrators report that they primarily encourage their teachers to use these types of assessments as well. Three-quarters of teachers reported that they definitely plan to use classroom-based assessments, and another 16 percent said that they will probably use them. In addition to classroom-based assessments, a majority of teachers definitely plan to use pre/post unit tests (64 percent) or assessments developed by their department, grade level or PLC (55 percent).

This finding stands in stark contrast to teacher’s views about using state tests as evidence of student growth, although the percentage of teachers who plan to use these types of tests has increased since last year. In 2013, less than one tenth of teachers (8 percent) indicated that they

definitely plan to use state standardized tests, compared with 79 percent of teachers who either might use (14 percent) or would not use them (65 percent). By 2014, seventeen percent of teachers reported that they would definitely use state standardized tests as evidence. A similar increase was seen in the use of district assessments such as MAP and DIBELS. Only 28% of teachers planned to use these tests as evidence in 2013, as compared to 42 percent of teachers who planned to do so in 2014. Upon closer examination of the 2014 data, a greater percentage of elementary school teachers (20 percent) indicated that they would use state tests, as compared to only 8 percent of high school teachers.

Generally speaking, teachers’ responses to this item in 2013 did not vary by district to any notable extent, with the exception of teachers’ planned use of district assessments. In two of the five districts, teachers reported that they definitely planned to use district assessments at nearly twice the rate of the three other districts. This likely reflected differences across districts in the extent to which these types of assessments were available and have been used in the past as part of the district’s assessment program. In 2014, some differences were also noted by school level with respect to the use of district assessments, with a higher percentage of elementary school teachers (64 percent) planning to use district assessments in comparison to 28 percent of middle school teachers and only 10 percent of high school teachers. Table 4 provides a display of teachers’ reported use of assessments as evidence of student growth in 2014.

Table 4: Types of Assessments Teachers Plan to Use as Evidence of Student Growth for Purposes of the Evaluation (n=1732)

	Definitely plan to use	Probably will use	Might use	Will not use/not applicable
Classroom-based assessments	75%	16%	4%	4%
Pre/post unit tests	64%	17%	9%	9%
Department, grade level or PLC developed assessments	55%	20%	11%	12%
District assessments (including assessments like MAP, DIBELS, etc.)	42%	13%	12%	32%
State benchmark data (e.g., math or reading)	23%	15%	18%	42%
Performance assessments (e.g., for art, music, PE)	19%	12%	11%	56%
State standardized tests (MSP, HSPE)	17%	10%	15%	56%

Dealing with assessments at the secondary level, and particularly in smaller districts may have its own set of challenges. A high school principal in one such context wasn’t encouraging his teachers to use any particular forms of evidence, as he thought teachers were able to figure out what to use in the initial implementation. “For right now, I think it’s getting us by and it’s got teachers in the right frame of mind.” In a small school where they don’t have many teachers

teaching common subjects, and few common assessments, the principal indicated teachers were talking about it and knew they needed to consider multiple forms of assessment.

Overall, the majority of school administrators did not encourage the use of state tests or state benchmark data. Sixty-six percent of school administrators did not encourage use of standardized state tests and fifty-three percent do not encourage the use of state benchmark data for any teachers. However, district views vary with regard to use of standardized state tests. One superintendent explained, we “can’t really use a [state] test measurement of student growth because it’s not designed to do that, secondly, it’s not a two-point in time measure... ‘what do we use to determine student growth’ is that really scary conversation for some teachers... Those are really different conversations and how does assessment play in to that, and how do we create assessments at the district level, at the school level?”

The 2013 survey also probed into the question of how teachers are organizing and keeping track of the evidence they are collecting. In this survey item, teachers were asked to identify all the ways in which they collect evidence. The largest proportion of teachers indicated that they were keeping hard copies of documents (79 percent). However, 60 percent reported that they were also keeping electronic records of student progress, and nearly the same proportion (58 percent) indicated that they were collecting evidence in their grade books. A growing number of teachers are using the eVAL tool, which will be discussed in greater detail in a subsequent section of this report.

Teacher challenges with goal setting and collection of evidence

In both the 2013 and 2014 surveys, teachers were asked to rate their level of confidence in setting goals and identifying evidence for student growth. The data indicate that a greater percentage of teachers felt “very confident” across all goal setting objectives in the second year of implementation. Overall, teachers felt more confident in setting goals for student growth than in identifying appropriate forms of evidence for student growth. In 2013, more than a third of teachers (37 percent) felt very confident in setting student growth goals in one of the three criterion areas, and by 2014 these percentages had increased by approximately 10 percentage points. More than 40 percent of teachers felt somewhat confident in these areas, and this percentage remained constant in both years. Most notably, only 28 percent of teachers felt very confident about identifying appropriate forms of evidence for student growth in 2013, but this proportion had increased to 42 percent in 2014. Table 5 details the changes in teacher confidence levels after one year.

Table 5: Teacher Confidence Regarding Goal Setting and Evidence of Student Growth (2013 n=1429, 2014 n=1732)

	Very confident		Somewhat confident		A little confident		Not confident	
	2013	2014	2013	2014	2013	2014	2013	2014
Set student growth goals for a whole classroom	37%	46%	45%	41%	14%	10%	4%	2%
Set student growth goals for a subgroup of students	36%	47%	42%	40%	17%	10%	4%	2%
Set student growth goals as part of a PLC, grade level or subject matter team	37%	47%	44%	39%	14%	10%	5%	3%
Identify appropriate forms of evidence for student growth	28%	42%	44%	41%	21%	13%	7%	4%

Subsequent analyses by district revealed that teachers’ confidence levels on all of these issues were substantially higher in one particular district, with more than half feeling very confident on setting goals and 48 percent feeling very confident about identifying forms of evidence for student growth. In this district, substantial professional development had been aimed at supporting teachers in learning how to develop growth goals.

Despite these higher levels of optimism, other findings suggest that teachers still have difficulty with the task of goal setting. For example, school and district administrators do not report the same levels of confidence in goal setting as teachers describe. As presented in Table 6, seventy-two percent of all principals and assistant principals, and 73 percent of district administrators rated teacher knowledge of goal setting as a “great” or “moderate” challenge in 2014. In four of the six districts, building administrators identified this as the greatest challenge facing their teachers with regard to the teacher evaluation. A second challenge widely agreed upon by school and district leaders is the ability of teachers to use formative or summative measures in developing goals. Lack of time was noted as a significant obstacle for teachers as well as building administrators, as district leaders identified the greatest challenge for principals as the lack of time for collaboration in setting goals for student growth.

Table 6: Perceived Goal Setting Challenges Faced By Teachers (Principals n=142, District Admin n=59)

	Not a challenge		A small challenge		A moderate challenge		A great challenge	
	Principal s	District Admin	Principal s	District Admin	Principal s	District Admin	Principal s	District Admin
Teacher knowledge about goal setting for student growth	2%	2%	25%	22%	54%	44%	18%	29%
Teacher ability to use measures of student growth in developing goals	7%	2%	28%	24%	48%	48%	17%	24%
Teacher time for collaboration with others in setting goals for student growth	11%	7%	31%	19%	31%	32%	27%	39%
Teacher ability to link student growth goals to improvement initiatives	8%	2%	35%	15%	39%	49%	18%	24%

Note: Discrepancies in percentages due to "don't know" responses

According to a district level educator, setting quality student growth goals has been a challenge for teachers in the district, in part because the district and schools did not provide teachers with examples of high-quality student growth goals. Another challenge was the extra work this task placed upon teachers, and differences in the messages that instructional coaches and principals were giving teachers about what constitutes quality evidence. As noted in Table 7, there is some discrepancy between teacher, principal, and district administrator beliefs with respect to the examples of evidence that can be used for measuring student growth. The greatest difference can be seen between teachers and district administrators, where only 19 percent of teachers feel they have received useful examples as compared to 27 percent of district staff.

Table 7: My district has provided examples of measures/evidence of student learning that teachers can use (Teacher n=1823, Principal n=142, District Admin n=59)

	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
Teachers	19%	45%	24%	11%
Principals	25%	45%	25%	6%
District Administrators	27%	46%	12%	8%

Despite the overall concern that professional development takes away from critical time spent in the classroom, over half of all teachers surveyed indicated that additional training would be useful in a variety of areas, including writing student growth goals (69 percent), identifying assessments that can be used to determine student growth (74 percent), and developing strategies to collect evidence of student learning (78 percent).

Need for appropriate formative assessments for teachers

Educators in the case study districts expressed a need for additional, appropriate assessments, which can support the integrity and reliability of the evaluation process. They also wanted more training in how to use these assessments to inform student growth goals. According to a district level educator tasked with helping to implement TPEP in the district, the most meaningful student achievement measures come from formative, classroom assessments. Thus she was grateful for the flexibility that the state evaluation has granted to districts in allowing them to choose their own measures of student growth. Still, she recognizes that her district is not yet to the point where evidence is consistent in quality across the grade-levels, subjects, and school levels. Another district administrator concludes that a central concern for the revised evaluation system is the need for appropriate assessments. She explained, “When we don't have a lot of assessments, typically what can happen is if you don't have a skill, the teacher is setting goals that are just not going to be powerful in terms of improving student learning... The limitation of assessments is going to be a huge challenge.” Even in a district that prides itself in being “data rich,” a district administrator explained that trying to work with assessment data in their PLCs can be challenging when trying to translate the assessments into instructional practice. Simply put, he said, “we don’t always know what to do [with the assessment data].”

For school and district leaders in the case study districts, authenticity of the assessments is of great concern. Several superintendents explained that addressing this challenge is one of the priorities of the current school year. One superintendent emphasized how he wants to create assessments that show genuine student growth, explaining, “I think there is a real way to play the game and this could be a lot of energy that doesn’t get us the kind of results I hope we can get and have more kids be successful.” His goal for the current academic year is to focus on how to measure student growth and how to create assessment instruments to do so. A middle school principal who is very interested in authentic engagement around TPEP described how he had this discussion with his staff:

“I want to make whatever we’re expecting to do, number 1, as natural a fit in what they’re doing as possible, and number two, as user-friendly as possible.... They are starting to talk about evidence, and evidence of student growth and they’re asking me, ‘what if I did this, what if I did that?’ I keep telling them that if this isn’t something that’s going to help you, if you’re just doing this for me as your supervisor, if you’re manufacturing evidence for that, that’s a waste of all our time. So trying to make it pretty authentic and embedded in what they would already do.”

Role of Collaboration

By design, the revised teacher and principal evaluation system was intended to broaden and deepen the conversation among educators about high quality instructional practice. The teacher and principal evaluation prompted conversations among district administrators, staff in human resources, building administrators, teachers, school boards (e.g., in the adoption of instructional frameworks) and teachers' and principals' associations. The early pilot and RIG efforts required memos of understanding between the teachers' association and the district, as these districts piloted aspects of the revised system with their teachers prior to the statewide implementation. In this section we highlight the various ways that educators interacted and collaborated around this work.

Broadened collaboration among multiple partners

From the beginning, TPEP had a significant impact on district-level collaboration. In the six case study districts, early conversations with teachers' associations proved critical to productive engagement around this issue. The case study districts sought to be intentional about including association representatives in discussions of how the implementation would play out, and the best ways to support staff through the process. Changes in association representatives over this time period, and sticky issues around contract language made things more difficult. Not all the districts experienced the same level of agreement in working with their associations. In one district, the TPEP process brought together a team of teachers and building level administrators, which the superintendent indicated was collaborative from the start, but unfortunately the teachers' association representative wasn't part of it. Later when it came to negotiation of the memo of understanding with the local association, there wasn't a clear understanding of the evaluation process. This delayed implementation until issues could be worked out through the collective bargaining process.

Despite these kinds of challenges, in several districts both district administrators and association representatives made a special point of mentioning how proud they were of the work they had been able to accomplish together in implementing TPEP. A principal in one district attributes the lack of tension with the teachers' association to the superintendent: "He's a very good communicator and he goes out of his way to explain things to make sure that the union understands what we're doing and why we're doing it. That this is to improve, not to punish... [he] does an excellent job of communicating with the community and with the teachers and the union." The director of teaching and learning in another district attributed the good relationship with the association to the fact that teachers have been involved with the process for a long time. TPEP wasn't new to them. "We haven't thrown this at them out of the blue. 'Hey! We are switching to [the framework] tomorrow and we've got a new evaluation system, and you fourteen people are going to be on it next year.' It just didn't happen that way.... I really think that involvement with the staff from the beginning and all the way through the process has

alleviated a lot of the fear and anxiety that has come along with moving all of these initiatives forward.”

A district administrator described the vision of the person responsible for TPEP implementation as key to keeping positive momentum with the process: “I think [her] vision, and this is, I don't want to try to necessarily try to label it individual, but I think [she] has a real authentic vision around collaboration. And she has really harvested a culture where teacher leaders come to the fore and grapple with complex problems. And she's patient. That has happened over time. I will be honest, when I first came on board, I've been in districts where to some degree that conversation hasn't been to the benefit of students. [She] really has pushed this group for that conversation not to be about the adults for the adults' sake, but about an adult need for the kids' sake. I've really watched her do that skillfully and that has been ongoing.”

Time and space for principal and teacher collaboration

At the building level, the places where collaboration has occurred around TPEP have been between the principal and individual teacher, and in groups of teachers working together. Professional learning communities (PLCs) and grade level or department teams of teachers can be natural places for teachers to work together on TPEP-related issues. Given the role that student growth criterion 8 (exhibiting collaborative and collegial practices focused on improving instructional practice and student learning) plays in prompting opportunities for collaborative learning, one might expect to see increased collaboration in these settings. Indeed, approximately half of teachers report a positive impact on the quality of professional collaboration (51 percent), and three-quarters of teachers indicate that conversations with other teachers about the progress of their students are useful in the TPEP process (76 percent). Most teachers report that their building has a collaborative school culture (42 percent strongly agree, and 42 percent somewhat agree), although elementary school teachers are significantly more likely to feel this way than middle school or high school teachers.⁴ Ninety-one percent of teachers surveyed report that they are part of a grade level, department or subject matter team, or PLC that meets on a regular basis. The majority of these teachers (58 percent) meet weekly, or every two weeks (19 percent), while the remainder meet once a month or every couple of months.

How well these spaces for professional community function, however, varied considerably by building, grade level and department, and whether or not these spaces are even used for activities related to TPEP. One district with a long history of PLCs as an organized space for teachers to come together to share goals and perceptions, also serves as a place to share with their association's representatives. An association representative explained that, “this district is big on PLCs,” and felt that these organized spaces were generally helpful for teachers as they adapted to the revised evaluation requirements. She reported, “We have some really strong PLCs and so I

⁴ $\chi^2(9, N=1705) = 49.85, p = .000, \text{Cramer's } V = .10$

think that the PLCs are now looking at implementing their focus on the focused evaluations, especially working together to do the same thing across the board.” Yet she also noted, “people who are not in a strong PLC, I think they are a little more lost. I think they are having a harder time with it.”

A superintendent in another district explained that there wasn’t necessarily increased collaboration among teachers because of TPEP if strong PLCs or subject departments were already in place. Rather he explained, “We are not collaborating more because of TPEP, but we are engaged with the instructional framework, and that is an advantage.” In other cases, districts chose to address TPEP issues separately. For example, one district chose to address TPEP issues primarily in late starts and release time, and less so in PLCs or grade level teams.

Several activities in teacher evaluation may be associated with collaborative efforts, including goal setting, collecting evidence of student growth, designing assessments, and analyzing student growth data together with colleagues. In the 2014 survey, we sought to understand whether teachers were participating in these activities primarily by themselves or with their colleagues. For example, Table 8 shows that over half of teachers reported working on goal setting for student growth (56 percent) and designing classroom-based assessments (60 percent) mostly with others. Conversely, a majority reported collecting evidence of student growth (77 percent) and analyzing student growth data (53 percent) mostly by themselves. We found variation on these items by district and also school level, as elementary school teachers were significantly more likely than high school teachers to work with others on goal setting,⁵ designing classroom-based assessments,⁶ and analyzing growth data,⁷ and middle school teachers were significantly more likely than high school teachers to work with others on collecting evidence.⁸

Table 8: 2014 Survey: Teacher Collaboration on Activities Related to Evaluation (n=1732)		
	Mostly with others	Mostly by myself
Goal setting for student growth	56%	44%
Collecting evidence of student growth	23%	77%
Designing classroom-based assessments	60%	40%
Analyzing student growth data	46%	53%

Of those teachers surveyed in 2014 who meet regularly with colleagues, we asked a series of questions about the extent to which they discussed effective instructional strategies in their PLCs, grade level or subject matter teams. As detailed in Table 9, the findings suggest that a majority of teachers worked together in discussing effective instructional strategies and

⁵ $\chi^2(3, N=1714) = 17.11, p = .001, \text{Cramer's } V = .10$

⁶ $\chi^2(3, N=1710) = 98.09, p = .000, \text{Cramer's } V = .24$

⁷ $\chi^2(3, N=1705) = 76.39, p = .000, \text{Cramer's } V = .21$

⁸ $\chi^2(3, N=1716) = 15.61, p = .001, \text{Cramer's } V = .10$

identifying assessments to measure student growth. A slightly smaller percentage of teachers collaborated on analyzing student work, writing student growth goals, and analyzing student growth data. Similar to the above findings on collaboration, there were significant differences between grade levels in the extent of collaboration within professional learning communities, departments, and subject matter teams. In particular, elementary teachers were significantly more likely than middle school and high school teachers to discuss effective instructional strategies together,⁹ and more likely than high school teachers to analyze student work,¹⁰ identify assessments for student growth,¹¹ and analyze student growth data together.¹²

Table 9: 2014 Survey: Extent of Teacher Collaboration on Activities Related to Evaluation with Grade Level, Department or PLC (n=1579)

	Very often	Somewhat often	Not often	Never
Discuss effective instructional strategies	36%	43%	18%	2%
Analyze student work together	15%	43%	34%	7%
Work together on writing student growth goals	20%	44%	31%	5%
Identify assessments that can be used to measure student growth	34%	47%	16%	3%
Collectively analyze student growth data	17%	39%	37%	7%

It should be acknowledged that we found some variation among districts on a few of these indicators, and that these represent, in some cases, differences in how they chose to approach these tasks as well as the implementation rollout (e.g., a majority of teachers on TPEP in the first year or staggered implementation).

In the 2013 survey data, we contrasted teachers’ perceptions with those of school administrators on several issues related to the collection of student evidence. According to school administrators, most teachers were working with their PLC or others in their building to collect and identify forms of evidence of student growth. This was highest in two of the three districts where the decision to work on goals was made collaboratively with other teachers at the school. Differences existed between the perceptions of teachers and school administrators about the extent to which administrators help teachers with the collection and identification of forms of evidence, with more than a third of school administrators (36 percent) stating they are involved to a great extent, as compared to only 14 percent of teachers. Table 10 compares teacher and school administrator views on these issues in 2013 (we did not ask the same questions in 2014).

⁹ $\chi^2(6, N=1539) = 38.65, p = .000, \text{Cramer's } V = .11$
¹⁰ $\chi^2(6, N=1532) = 81.55, p = .000, \text{Cramer's } V = .16$
¹¹ $\chi^2(6, N=1534) = 33.61, p = .000, \text{Cramer's } V = .11$
¹² $\chi^2(6, N=1533) = 103.53, p = .000, \text{Cramer's } V = .18$

Table 10: 2013 Survey: Teacher and Principal Views on the Collection of Evidence of Student Growth (Teacher n=1429, Principal n=129)

	To a great extent		To some extent		A little		Not at all	
	Teachers	Principals	Teachers	Principals	Teachers	Principals	Teachers	Principals
Principal or assistant principal helps with collection and identification of forms of evidence	14%	36%	34%	43%	27%	19%	25%	3%
Teacher is working with PLC or others in building to collect and identify forms of evidence	34%	44%	38%	39%	18%	14%	9%	3%
Teacher collects and identifies forms of evidence by him/herself	57%	21%	31%	46%	9%	27%	2%	5%

The collaborative work around student growth goals and the collection of evidence has proved challenging, even in these six districts where effort has been made to support these activities. An elementary principal explained how student growth goals in her school have been based on the school improvement plan, but collecting evidence of student growth proved challenging, even with talented and veteran staff. Her teachers were not used to looking at data and figuring out how their instruction should change or what they should do differently throughout the year. But she says, they are getting better at it. She had to create space for them to do this by setting up periodic opportunities for teachers to discuss student grade level data together. The principal described how she carefully structures those meetings so that everyone brings the most recent data collected around student growth. Further, she also organizes the teachers in mixed groups with different grade levels making clear, “When you have to explain it to another grade level you analyze it a little bit better. Teachers can see what’s happening above and below them.”

Challenges for inclusion and integration of teachers in specific roles

The TPEP process pushes professional learning by exploring what instructional practices may be lacking within school settings, and what supports might be needed for teachers. It also reveals that teachers in particular roles (e.g., special education, PE, art, and other specialists) may benefit from greater integration and inclusion in professional learning communities. Teachers in certain subject areas or specialist assignments indicate that the evaluation doesn’t always neatly fit their work. What may be an appropriate instructional strategy for teachers in some disciplinary fields may not work well or even be appropriate for those teaching in certain performance fields or with particular subpopulations of students. In some cases, principals or other evaluators may not understand the rationale or be able to accommodate the differences, which could result in teachers being marked down. A music teacher in an urban district explained, “So many times we have to explain, and really teach our principals the music EALRs...what it means to teach music. For example, my principal challenged me for teaching the same lesson to one grade level as another. We do this intentionally. Of course our objectives are differentiated by grade levels, but the lesson is still on an element of music.”

Some of the case study districts recognized the need for specialists to collaborate around TPEP implementation issues. In one district, specific committees were created for job-alike specialists across the district to meet on a regular basis to share their work with goal setting and the collection of evidence. In some case study schools, PLCs were organized in such a way as to facilitate collaboration among specialists so they could obtain necessary supports.

Increased teacher voice, engagement and responsibility for the evaluation

One of the ways the revised teacher evaluation system has increased collaboration is between the teacher and the principal or assistant principal who is evaluating the person. Indeed, many principals and teachers participating in the study affirmed the richness of conversations about instructional practices, and considered this aspect to be among the most beneficial of the entire process. Beyond that, it can provide teachers with the opportunity to be actively involved in their evaluation by contributing evidence of their teaching practices, and giving them a stronger voice in the outcome. A high school principal explained from his perspective this benefit:

“The really nice thing about [the framework] is that it gives teachers a voice.... I show them where I think they fit on the [framework] model, and they tell me what they think... and 96% of the time [we agree] and then that 4% of the time where we don't agree, they are given the offer to come back with evidence, and they do. The beauty of that is they'll show me all sorts of things that are occurring, journal writing, responses in journals, back and forth dialogues, emails, things that I wouldn't have access to and could never see just sitting in their room. So it gives them that voice, it gives them that opportunity to come share all those things with me. To me that's hands down the best piece of all of this. We've always had those dialogues, we've always had those conversations, but I've always been the one who just checked the box and they signed off. Here they have a chance to check their own boxes and really have a chance to score themselves.”

Taken together, TPEP implementation has prompted numerous forms of collaboration at multiple levels of the education system. In some cases, a strong foundation and history of prior collaborative work has facilitated the process. In other districts and schools, the need for increased collaboration has been realized. Challenges do exist with these forms of collaboration, particularly for teachers who work in specialized roles.

Evolving Professional Development

After the first 18 months of statewide implementation, there is now greater awareness of the TPEP process among school and district staff. OSPI estimates that virtually all principals and 70% of all teachers statewide will have experienced an evaluation under TPEP by the end of the

second year (personal communication, 2014). Given this fundamentally different evaluation model, our findings suggest that professional development is needed across all instructional and administrative staff.

Initial approach and the need to develop local capacity

As pilot or RIG districts, the six case study districts had a head start in introducing their staff to the revised evaluation system. The state provided some materials and training for professional development, and the districts used the grant funding they received for release time for training and collaboration, to purchase training from framework and technology specialists, and work with instructional coaches.

Most of the districts provided staff with paid professional development days in the summer and substitutes during the year when needed. They brought in framework trainers or specialists to present the instructional model, highlight vocabulary, watch videos on scripting lessons, and have their staff complete self-assessments. Through a combination of regular professional development days and grade level or content level team meetings during the year, districts sought to create time for professional dialogue. Several of the districts used a “train the trainer” model to present the instructional framework and provide support to learn it. In some cases, the Educational Service Districts served to help facilitate communication about the new frameworks and evaluation.

While the standard statewide training available to districts has been widely used, some of the case study districts are now focusing on their own locally developed and embedded training. Decisions about the content and format of professional training are being driven partially by local concerns and needs of teachers. In some cases there has been a shift to working with novice teachers or staff new to the district, who may not be familiar with the framework. The use of outside providers of professional development has fallen out of favor in some cases. A superintendent explained that after bringing in expertise from the outside, his teachers now want to see more local practitioners in their classrooms who can provide relevant experience and examples, and less of a theoretical orientation. In another district, an administrator explained that his staff now need to sit down with the framework and analyze the skills they need to be successful. He worried out loud about whether teachers have been trained to work successfully in PLCs, and to write effective formative assessments.

Teaching involves a very complex set of tasks which the instructional frameworks have attempted to conceptualize and organize. In practice, the frameworks also identify areas, such as assessment literacy, where many educators have gaps in their knowledge. Formative assessment, in particular, is essential for this evaluation system to work. Teachers may understand the student growth requirement, but they may not know how to choose the right type of assessment to demonstrate student growth, or to match the assessment to the needs of their students. In the

2014 survey, nearly three-quarters of teachers (74 percent) identified “training in assessments that can be used to determine student growth” as a topic that would be useful to them, and 78 percent identified the topic “strategies for collecting different forms of evidence of student learning,” as an area of training they need.

One district found that early professional development activities weren’t very productive. They have since directed staff to training which is grounded in formative assessments, to be used as a jumping off point for their PLC work. Another district admitted that they couldn’t find anyone to do the assessment session for their district training, and they had to rethink how to provide this training for their staff. Other districts have recognized the need to continually cycle back through certain professional development topics for staff who are new to a school or district.

Professional development support for teachers

Districts have considered a variety of ways to provide professional development support to their teachers. As mentioned earlier, some schools and districts have introduced aspects of the framework or related instructional practices incrementally, or by starting with less complicated components first. The timing and integration of professional development within PLCs or grade level or subject area groups is another approach that some have used. Some districts provided differential supports for new and experienced teachers. For example, in a district that describes itself as a “transitioning district,” where new teachers start out before moving elsewhere, administrators described designing meetings with new teachers specifically to discuss TPEP. Other districts also used this approach as part of their induction program for new teachers.

In addition, many schools have increasingly relied on instructional coaches. A teacher leader and coach explained, “I think that as teachers become more fluent with [the framework] and principals and [instructional coaches] work through those feedback conversations, I think those instructional coaches need to be part of ‘what next?’ You know, as we sit down, and you and I define what my next big step is for growth, then that instructional coach needs to be a resource that is really brought right into [the process] more formally...” While this strategy has some advantages, such as lessening the workload of principals and allowing teachers to learn from teachers, one principal felt that as a result, he didn’t have a good grasp of the model himself. He reported that he is also receiving the training now, which has been helpful.

In the surveys, teachers were asked about the extent to which their district had provided training on goal setting for student growth. Overall, 20 percent of the teachers surveyed in 2014 “strongly agreed” and 50 percent “somewhat agreed” that the district provided training on goal setting that had been useful, indicating a total of 70 percent of teachers felt positively about the training they had received. However, responses to this item varied significantly by district, ranging from a low of 39 percent in one district to a high of 87 percent in another.

Professional development support for principals

Building administrators carry much of the load in the implementation of TPEP. They have needed substantial support in learning the evaluation model, especially as changes have occurred since the initial piloting process. As mentioned earlier, some districts have chosen to focus their training on certain framework components within or across several years. A district administrator explained, “The problem with all the frameworks that are being used in the state is that they are so big, and there’s so much information.... Even those of us in education, we aren’t the experts on instruction that we sometimes think we are. We know pieces of it. But now we are going to put it all together. And we thought a way to do that would be to bite off a chunk of the pie each time. So [in] each meeting we do one of the components of the Danielson framework.”

All of the case study districts utilized some form of leadership development support for their building administrators. This often took the form of regular weekly or monthly meetings in which TPEP was a major focus for the districts. One district formed a “critical friends” model of support for principals. In another district, building administrators come together two mornings a month in leadership development sessions to work on issues related to TPEP, Common Core and their own PLCs. In the past several years, they spent at least half of their time on TPEP, discussing issues such as rater agreement, watching videos and having discussions on what evidence is and is not. These meetings also served the purpose of helping to develop a common language among the different building administrators. On the 2014 surveys, 42 percent of principals and assistant principals indicated that the training they received in implementing the revised evaluation system was very useful (37 percent somewhat useful), and 47 percent found time to work with other administrators on TPEP issues to be very useful (31 percent somewhat useful).

As part of their training, but also to address concerns about rater agreement across school sites, building administrators in several districts regularly participated in “rounds” to schools to observe teaching, and later to discuss and try to come closer to agreement about what good instruction should look like. In another district, all building administrators were watching videotapes in Teachscape and applying the rubrics to the lesson observed. A district administrator indicated that the principals really like Teachscape and share it with their teachers, “because we can see it, it’s more than reading about it on paper.” As a result of teachers’ interest, this district has also provided opportunities for teachers to complete training using Teachscape. The issue of rater agreement and how to ensure that principals implement TPEP consistently across schools was a theme expressed by numerous individuals and levels of the school system. Several districts sought to address this issue by holding summer in-service training for administrators. Some principals also have completed training through the Association of Washington School Principals to supplement their understanding of the teacher and principal evaluation.

Restructuring Human Resources

Arguably the greatest challenges to effectively implementing TPEP are time and human resources. The workload for building administrators can be overwhelming. Some Washington educators estimate that principals will realistically spend 10 to 14 hours per teacher in conducting a comprehensive evaluation, and that is on top of running a school. An equally daunting task is managing the financial costs to support the implementation of both the revised evaluation system and the Common Core State Standards (CCSS), with the need for additional staff and professional development support. As a result, all six of the case study districts restructured human resources to accommodate the increased workload.

Differences between elementary and secondary schools

In the case study districts, TPEP implementation differentially impacted elementary and secondary school administrators. It has been suggested that TPEP implementation is more challenging at the elementary level because of workload on the principal. Elementary teachers typically teach every subject, and the elementary principal is often expected to be the instructional leader in every subject at that school. It is not uncommon for a principal to be responsible for evaluating 30 teachers, in addition to evaluating the non-instructional staff in the building. At the secondary level, while there are often more teachers, there are typically more administrative staff to share the responsibilities. In addition, secondary principals often have other administrators in the building with whom they can collaborate and discuss TPEP issues, which is not necessarily the case for elementary principals. While it depends in part on school size, the elementary principal is often the only administrator in the building.

However, differential impact on administrators isn't always related to the building level. One district administrator explained that differences in implementation may be due to principal leadership: "how far this is going with staff is directly proportional to the competency or how focused the principal is...It's hard to lead something you don't totally grasp yourself." The experience, management or leadership style of the building administrator also influences the challenges they face. Additionally, the extent to which teacher leaders (e.g., department chairs, instructional coaches, association representatives) are available and involved with TPEP implementation also impacts the principal's scope of TPEP-related activities, such as organizing and leading professional development.

Finally, some principals have developed strategies to maximize their time in conversation with teachers about elements of TPEP implementation. The principal of a large comprehensive high school describes how he maximizes time with teachers who are on a focused evaluation plan: "Essentially what I am doing is spending time with them in their PLC groups in our early release days and using that as my evaluation. Then we meet separately and talk about the student

growth. The student growth is all correlated to their PLC goals...so that frees me up to spend time with my nine [teachers on comprehensive].” He also indicated that small group instruction is a good way to communicate information to staff.

Supports for principals in their buildings

A high priority for district administrators in all six case study districts was to provide additional support for administrative staff who were responsible for evaluating teachers. But over a third of principals (35 percent) surveyed in 2014 disagreed that their district provided the staffing support that they need to effectively evaluate teachers. Supporting principals is complicated by the nature of their work and by virtue of the fact that some school leaders struggle with delegating aspects of what they perceive to be core responsibilities to others in the building. In a later section we discuss the changing role of the principal. In this section we focus on supports that districts have provided for administrative staff.

Adding assistant principals or other similar staff. All six case study districts added part-time or full-time assistant principals to some or all of their elementary schools during this time. In one of the smaller districts, 1.5 assistant principals were added at the elementary level, split between 3 schools. Another district strategically added the assistant principals at the elementary level to their highest poverty schools. At the secondary level, a larger district organized the high schools to have a dean of students and an athletic coordinator to free up the principal to focus on supervision and professional development. Another district hired deans of students for their most highly impacted secondary schools.

TPEP has also prompted some assistant principals to take on a greater role in instructional leadership, in support of the teacher evaluation effort. For example in one middle school, a staff member moved from a position in athletics to instructional leadership as an assistant principal. The 2014 survey data confirm that for principals who were provided additional administrative staff in their building, 73 percent found this support to be very useful (14 percent somewhat useful).

Adding instructional coaches. Many of the case study schools were already staffed with an instructional coach. Some of these coaches were designated to support the TPEP implementation. In one district, all elementary schools received a designated instructional coach, while the secondary schools shared instructional coaches in math and language arts. The 2014 survey data revealed that more than three-quarters of principals (79 percent) who were provided an instructional coach in their building, found this support to be either very useful (61 percent or somewhat useful 16 percent).

Training instructional paraeducators. In one district, instructional para-educators have been included in some of the TPEP training. The principal explained that paraeducators have been part of the

process of understanding what good instruction looks like. Along with his secretarial staff, the principal used a cycle of inquiry to focus on things he wants them to do, including evidence of how well it is working.

District level changes in staffing

In addition to building level changes in staffing, all six case study districts made staffing changes by adding or re-allocating positions at the district level to support instructional improvement initiatives. One district reassigned a former principal to a new district-level teaching and learning position. In this new role, the administrator now directly supports principals and leadership teams in literacy and math, as well as evaluating a few teachers and some principals. Another district also moved a former principal to a district-level support position to assist principals specifically around TPEP issues.

In discussions with districts around the state about support for administrators, the conversation included potential use of external evaluators for the TPEP process. While some administrators admit it could potentially reduce bias in staff evaluations, others argue that a more holistic evaluation can be conducted by a building administrator who has a relationship with the staff. Unless the evaluator has a longer term instructional role like a coach, the outcomes for teacher improvement are harder to support. Additionally an external evaluator would not see all the different facets of a teacher's work that happen naturally in a school, from their meetings with parents to other classroom activities which are typically observed by a principal. Another administrator explained, "I have a hard time giving away the comprehensive evaluations. I think the principal needs to do that. I think where we can get some support for them on is on the focused evaluation."

New hiring practices

Nearly all the case study districts also described changes in hiring practices as a result of TPEP implementation. A district administrator explained that as a district, they were a lot more cautious: "...[we are] looking at their experiences and background to make sure that they meld with what we're doing here." He indicated that when people asked about how to get hired in the district, he would tell them that they should understand the instructional framework and Common Core State Standards. During the interview process in another district, potential teachers were asked to research TPEP and present on the TPEP standards as part of the teaching lessons during the interview process. In this way, the district has re-written their entire hiring process to focus upon TPEP so that the reference checks and interview questions align with the framework. Another district also mentioned that they changed the job description of some staff to align with the requirements of the teacher evaluation.

Shift rather than decrease in the workload over time

Given the enormous effort needed to implement the revised evaluation system, some had hoped that after an initial period of adjustment, some of the workload issues might settle out or even decrease over time. However, in discussions with staff in the case study districts, they reported that the workload has not decreased. Rather, as districts have been at it longer, the work has shifted. For districts that opted to stagger implementation, their workload will likely increase in the third year, as the remaining teachers are added to the revised evaluation system. But even for the districts that transitioned all staff to TPEP from the beginning, many indicated that the workload has not diminished by the second year, but instead has changed.

An elementary principal explained that things don't go away, but she finds that she is more intentional in prioritizing and delegating some tasks. Several principals talked about using principal interns, or teachers working on administrative credentials, to support discipline and management in the school or using teacher coaches in ways that free up the principal to address other priorities. One principal indicated that in her district, even prior to the revised evaluation system, principals had to be in classrooms two hours a day by contract; now principals use the time in classrooms for TPEP matters. However, numerous principals indicated that the management of the paperwork has increased from prior years.

But not everyone agrees that workload is an issue. In response to a question about the increased workload, a high school principal responded: "It increases the workload in that group of comprehensive folks that you're working with. It is an increased workload there. But is that a meaningful workload? Is that the heart and soul of what we're supposed to be here to do to improve instruction and practice? Do you see yourself as an instructional leader? Then the template is there." The following section of this report continues with a discussion of the leadership issues associated with TPEP implementation and how these issues are understood by various stakeholders.

Role of Leadership

Most educational leaders in Washington state would agree that teacher evaluation needed an overhaul. The old system rarely helped teachers improve nor did it distinguish between those who were highly effective in specific areas and those who were not. The school and district leaders in the six case study districts approached the enormous system-level changes needed to implement a revised evaluation system by promoting a conception of teaching and learning that supported continuous improvement. These leaders took the long view of implementation for comprehensive change within their local contexts, and expected that it would take a number of years for the evaluation system to be operating at an optimal level.

A key finding from this study is that district and school leadership actions set the tone for productive engagement with staff in the implementation efforts. As early implementers of TPEP, the message across all six case study districts was one of using the evaluation as an opportunity for instructional improvement and conversations about effective teaching and student learning. Leaders in the case study districts directly connected the evaluation system to the instructional improvement agenda, and adopted a learning stance towards teacher evaluation, emphasizing the potential benefits rather than viewing the revised system as yet another state mandate requiring compliance. They were also protective of their staff, recognizing that they could not hold them accountable if they had not provided appropriate supports.

Finding the right balance

The school and district leaders in these districts sought to alleviate concerns by directly involving staff in the teacher evaluation process. A theory of action guiding the adoption of TPEP was that it is no longer solely the administrator's job to evaluate teachers. Rather, according to one superintendent, it is "a partnership where the teacher brings evidence to the table and administrators gather evidence in other ways." A high school principal shared that when they started to prepare for TPEP implementation, the first thing they tried to do was alleviate concerns and clarify expectations. He explained, "Show them how simple this is. How there are 8 criteria, how there's student growth. [The framework] works well with that...." Overall, these districts tried to get ahead of the curve with messaging about the purpose of the evaluation and how their district had chosen to approach it. According to the 2014 survey data, 65 percent of teachers agreed, either somewhat or strongly, that their district provided clarity about the expectations for the evaluation process.

In one district, the superintendent explained the emphasis on doing what is supportive for their staff, while operating within the realm of the evaluation system. This meant they weren't overly concerned with all the implementation details in the initial stages. This district embraced the opportunity, and especially the use of the framework, to promote professional growth and focus on good classroom practice. In some ways, the six case study districts can be located on a continuum from those with a more relaxed approach to implementation to those that attempted to do it all from the beginning. In doing so, a pattern emerges as to whether there was enough structure to support activities without creating too much pressure that negatively impacted staff. In other words, where the expectations for certain levels of performance were very high, teachers and administrators expressed more reservations about the evaluation system. The notion of staff-buy in was related to both district and school leadership and how well these leaders were able to focus the initial evaluation process in ways that were manageable for staff.

Expanding role of the principal

Changes in teacher evaluation illustrate the expanding role of the principalship. While this notion has been discussed in recent years, teacher evaluation brings it into sharp focus because it

touches on so many aspects of instructional practice. Educators are again considering whether one person in the role of principal can be expected to carry out all the responsibilities associated with being both an instructional leader and a building manager.

Some educators among the case study districts have suggested that it is time to look at the principalship differently. For some school leaders, changes to the evaluation system without sufficient supports in place have taken a personal toll. A principal at a middle school expressed that the first year of statewide implementation was very difficult with both TPEP and the new requirements posed by the Common Core State Standards, and that the workload was not sustainable. He explained, “I just can’t do the hours I did last year. We worked almost every Saturday, every Sunday night just trying to keep up. The workload has changed. The job has changed.” The superintendent in this district described his principals as trying to find a rhythm in the work. He recognized that no two principals approach the evaluation work in the same way, and that principals are in different places in terms of their ability to do the work.

One of the districts has begun a conversation about the role of the principal, and whether current responsibilities have created a nearly impossible job. As the case study districts have done throughout the initial implementation of TPEP, they are now on the leading edge of trying to figure out appropriate expectations for the role of a principal. A novice middle school principal articulated the dilemma in the following way: “What this evaluation system has done is really put the demands of this job in a totally different context. In a good way, I think... And if you’re not pretty excited about instruction and improvement... then gosh, this is not very much fun.”

Identifying, distributing and supporting instructional leadership

The revised teacher evaluation system places a sharp focus on the instructional leadership capacity of the principal. This requires a comfort level on the part of principals to delegate other responsibilities that they have typically assumed for themselves. This shift in priorities can be difficult for principals, especially when parents and members of the public (and to some extent school staff) often hold certain expectations regarding what school principals should be doing.

One district that used a “menu of supports for principals” found that “very few of those resources have been accessed by principals.” The superintendent continued, “Which raises for me the question of, if our principals value the TPEP process, and I hear over and over again that they do, and I don’t think anyone is saying let’s go back to what we had before.... I think the challenge that we have fundamentally is that our principals may not be willing to let go of anything. And so to use the resources in the form that resources are available, is to in effect delegate or subcontract some portion of their work, and they are reluctant to do that.... It’s hard for them to let go of the portion of the work that they own deeply, that they are responsible for everything that goes on in their school.”

As mentioned earlier, some school leaders have discovered ways to use teacher leaders, coaches and others to build staff capacity. Leaders acknowledge the strengths of their staff and find ways to capitalize upon them. Additionally, some principals are learning how to share leadership responsibilities in ways that further support or deepen the learning improvement agenda for the school, thereby creating coherence between the work of TPEP and the work of improving teaching and learning.

Connections to Common Core State Standards and other initiatives

One of Linda Darling-Hammond's (2013) five key elements for a teacher evaluation system is adoption of common statewide teaching standards that are aligned with student learning objectives. Washington state has adopted the Common Core State Standards (CCSS) and the question now becomes, to what extent are goals for student growth in TPEP and CCSS aligned? Fifty-seven percent of teachers and 56 percent of principals surveyed in 2014, agreed either somewhat or strongly, that their district provided clarity about the connection between the teacher evaluation and district-adopted learning standards. Among district administrators surveyed, 63 percent indicated that information was provided on how the teacher evaluation is connected to the Common Core.

Some of the case study districts indicated that they have sought to integrate these two initiatives and that the staff are making connections between them. One superintendent explained, "I used to hear about how disjointed the initiatives were, that there were too many parts to pay attention to, to learn. People haven't been saying that as much." He attributes this to principals seeing the connections and communicating that better. For this superintendent, "the [framework] is the how, and Common Core is the what."

However, introducing a revised teacher and principal evaluation system concurrently with other major statewide initiatives, such as the Common Core State Standards and new assessments under the Smarter Balanced Assessment Consortium (SBAC) has been challenging for many districts. In addition, districts have other local initiatives underway to improve teaching and learning. For some districts, while the initiatives were implemented at the same time, the work has been compartmentalized among the specialized district staff who hold responsibilities for specific initiatives. A district administrator explained that messages were getting crossed, creating confusion for staff. As he sees it, district-level staff may need to give up some turf and become more intentional and unified so that educators can make the linkages between the improvement of instruction, the revised evaluation system and the Common Core State Standards.

With multiple initiatives in place in another district, an administrator indicated that they have tried to show the connectedness of everything they do: "All of our ELA staff has really come to the realization that their instruction has to change. The nice thing about that is that we can then

always tie it back to the instructional framework. And I'm always saying, 'Isn't it nice that we're talking about instruction while we're doing this' ...And so, just tying it back all the time, and making sure that they're not getting overwhelmed because it is overwhelming when you talk about TPEP and you talk about Common Core and you talk about the principal's evaluation, and everything that we have coming on board."

A district administrator described the work of connecting TPEP with the Common Core State Standards as "a lynch pin idea." She continued:

"I think if we don't think about how it's connected it will be fragmented and seen as being outside of the core work we do every day. So, one of the things I've been pressing [the department] to do, is what I call 'weave.' So, for example, Common Core, when we're doing Common Core, what are the connections and the through lines we can run through the teacher evaluation and the principal evaluation so that principals and teachers see this as a part of the work that they are doing?What we'll actually be doing is helping teachers to get better at their craft, helping principals to be better at observation and feedback and pivoting back out to the framework to use that language to have those conversations with teachers. And while they're doing that, there are going to be places within the framework that talk about teaching the standards. Oh, which standards? Common Core State Standards. So then that connection is natural and organic. To me, the most important part of that has to be that the leaders see those connections themselves and they are making those natural connections work for staff."

Given the numerous tasks involved and the related workload associated with TPEP implementation, tools have been created to help provide support for the evaluation process. The following section describes how electronic tools are being utilized by administrators and teachers in the case study districts.

Use of Electronic Tools

Technological tools have played a role in the initial implementation of TPEP. In particular, eVAL, a locally developed web-based tool has been used to support the TPEP implementation. eVAL was designed as part of a collaborative effort between the Washington Education Association (WEA), the Office of the Superintendent of Public Instruction (OSPI) and Educational Service District 113. It is offered at no cost to Washington school districts, primarily to support principals in supervising and providing feedback to staff throughout the evaluation process. According to data provided by the developers of eVAL, 153 districts across Washington have tried using eVAL since its introduction, and 105 are currently utilizing it in the second year of implementation. How the eVAL tool is used varies across these districts, with

some using eVAL to support observations, conduct self-assessments and complete scoring, and others using it primarily to store artifacts and evidence.

Variation across and within districts in the use of technological tools

Among the six case study districts, use of eVAL or other technological tools also varies widely. Teachers and principals in Evergreen, Kelso, Mount Vernon and Cashmere have used eVAL in a variety of ways. Renton chose to develop and create their own electronic tool called eWALK, and Kennewick has opted to use other electronic and paper tools at this time. Based on survey and interview data, some teachers and principals welcome eVAL as a tool in the evaluation process. Some principals like having an electronic tool to help them manage the evaluation process, and some teachers like the idea of having something to organize their evidence.

In the four case study districts where eVAL is commonly used, 2013 survey data indicated that one third of teachers and 53 percent of building administrators had logged on and tried to use the eVAL tool. By the fall of 2014, 56 percent of teachers and 70 percent of building administrators in these districts indicated they would be using eVAL to support their evaluations. More specifically, of those who were using eVAL in 2014, 75 percent of teachers reported using the system to share information with their administrators, and 56 percent reported receiving principal feedback through eVAL. Not surprisingly, the extent to which teachers shared documents on eVAL was significantly related to the extent to which principals sent feedback to them using eVAL.¹³ As one teacher described, he has used eVAL to organize his evaluation documents, but “To my great disappointment, my administrator does not provide any feedback throughout the year. After a year and a half, I have given up checking the eVAL system for feedback.”

Rates of eVAL utilization by teachers differed across several variables. In the 2013 survey, high school teachers were found to use the eVAL tool at higher rates (54 percent), than middle school (38 percent) or elementary school teachers (17 percent). Larger proportions of teachers reported using eVAL in 2014, and comparable to 2013, high school teachers were significantly more likely to use eVAL (86 percent), as compared to 54 percent of middle school teachers and 45 percent of elementary teachers.¹⁴ Differences in eVAL usage also existed based on years of teaching experience, where those who had been teaching for less than 4 years were significantly more likely to use eVAL (66 percent) in comparison to those who had been teaching for 4 to 10 years (61 percent), and for over ten years (53 percent).¹⁵ Differences in teacher eVAL usage by school district varied from 42 to 87 percent.

But as with early versions of most new software programs, eVAL experienced early bumps in usability which required multiple changes to the system. In the first year of implementation,

¹³Spearman's $r_s = .421, p = .000$

¹⁴ $\chi^2(4, N=1058) = 126.38, p = .000, \text{Cramer's } V = .244$

¹⁵ $\chi^2(4, N=1054) = 14.99, p = .005, \text{Cramer's } V = .084$

principals frequently assumed data management responsibilities in eVAL for their teachers, in an effort to ease the burden of this aspect of the evaluation process. An elementary principal in one district opted for a different solution. Because fully implementing TPEP was very complex and time consuming, the principal felt using eVAL would have added another learning curve for her staff. Instead, she designed an Excel spreadsheet which summarized the framework domains, criteria and scoring rubric in a simplified way. In evaluating her teachers, she explains that she takes literal observation notes in a Word document, and she sends those in an email with notes to the teacher. Later she categorizes the data according to domain and works through the rubric all within one page of an Excel spreadsheet. She believes this fulfills the level of information that her teachers need and the requirements of the evaluation without complicating the process.

Three of the case study districts have strongly encouraged all their principals to adopt eVAL as a tool for the evaluation process. One of the districts left the decision largely in the hands of principals as to whether or not eVAL would be used with their staff. The superintendent in this district explained, “We allowed people who wanted to use it to do it, and we have a dozen principals who used it with some teachers, [others] not at all. We have a couple of principals who used it with everyone and honestly we find it, as soon as you crack the code, and the code’s not that hard to crack, but all of us have a learning curve around that, we are finding that it’s a really helpful tool.”

In cases where it has been an effective tool and the staff use it, it is often because the principal is comfortable with technology tools and teachers are willing try the technology. A teacher involved in the TPEP rollout described herself as being someone who enjoys technology and she is “a big fan of it.” However, her principal, who is less proficient with technological tools, struggled to use the tool despite investing much time to seek out training. Part of the problem was that the tool was updated, which altered the format and functionality of the tool in ways that made it frustrating for her principal, who ultimately decided to forgo the tool and return to a “paper-pencil form.” However, the same teacher provided an example of how a “very tech savvy” principal took to the program easily and used it in the evaluations. She explained, “Right now, I think it is going to be the principals that are tech savvy and they will kind of go through and find the bugs and figure out how to do it.”

An elementary principal who was comfortable with technology tried to conduct many of the pre and post conferences with his teachers on eVAL. But the teachers found it too impersonal and wanted to have a dialog in person, rather than an email that said ‘Check eVAL for my response and questions to you.’ The school leadership team pulled the principal back a bit to help him work through the issues: “I was very gung-ho thinking we were going to use it, and then they did a great job of kind of reeling me back to earth.”

Renton developed their own electronic tool called eWALK. Among Renton teachers responding to the 2014 survey, 94 percent indicated they were using eWALK to support their evaluation this year. eWALK, like eVAL has faced some early technological challenges, according to teachers. Renton teachers' use of eWALK is higher than teachers' use of eVAL in other districts. Specifically, 87 percent of Renton teachers use eWALK to share documents with their administrator as opposed to 75 percent in other districts, and 72 percent of Renton teachers receive feedback through this electronic system from their principal as opposed to 56 percent in other districts. Additionally, district administrators have been able to use eWALK to identify areas where teachers may need additional support and to target professional development in those areas.

Views of the benefits and challenges of eVAL

Teachers who have tried to use the tool are fairly evenly split regarding their views of whether eVAL is relatively easy to use or whether eVAL helps them manage the evaluation process. What is clear is that teachers and administrators are becoming more familiar with the eVAL tool. In 2014, 64 percent of teachers surveyed agreed (either somewhat or strongly) that the eVAL tool was relatively easy to use compared with 55 percent in the prior year. A higher percentage of teachers in some districts found it easy to use eVAL as compared to those in other districts (70 percent in one district versus 48 percent in another). Closer examination of the survey data revealed that the perception that "eVAL is easy to use" is significantly positively correlated with the professional development and support they received to use eVAL.¹⁶ Teachers' open-ended comments on the 2013 mid-year survey about eVAL supported this conclusion, as a majority of open-ended comments from teachers and administrators referenced the need for more technical support and training.

Open-ended survey responses from the fall of 2014 revealed similar findings. A few teachers and administrators expressed satisfaction with the tool, but the vast majority of responses spoke to the need for eVAL training. As one teacher stressed, "Our administrators want us to use this system, but absolutely no training has been provided. As teachers we do not have the time to teach ourselves how to use this system." From the administrative perspective, a school principal commented that "It would be helpful if there was training for administrators about how to effectively use eVAL to do our jobs." It is likely that most people who felt comfortable with the software did not offer comment on this topic, although one person shared that "There have been more trainings available on an optional basis that I have chosen not to attend because I feel relatively confident navigating through the eVAL system and completing TPEP. When I am confused or do have questions my administrators are very helpful and can clarify things for me." The availability of this type of support seemed to vary dramatically by district as well as school.

¹⁶ Spearman's $r_s = .312, p = .000$

Other open-ended comments from 2013 and 2014 included themes such as technical difficulties with the software (e.g., login issues, lack of browser compatibility, timing out, and lack of intuitiveness), ways people have used it (to complete the self-assessment and for uploading artifacts), and the belief that using eVAL is too time consuming. One teacher shared that “While I use eVAL to complete my evaluations, it is sometimes not intuitive to use and requires a lot of unnecessary steps to enter comments.” Another teacher indicated that “The eVAL system is great but it is hard to upload. My principal could not see my evidence. If I hadn't backed up my work on a USB drive all would be lost.” One suggestion offered was that “making the eVAL tool teacher-centered rather than task-centered would be a huge support. Right now, I have to identify the teacher I am working on and then re-identify them in each of the task areas. It would be better if I selected the teacher, and then had all of the tasks available for that teacher, without having to re-select them each time.” In summary, many respondents seemed interested in using eVAL, but hoped for more professional development and continued improvement of the tool.

Among the building administrators who utilized eVAL in 2014, 70 percent use it to share information or documents with their teachers, and 93 percent are using it to create a final summary record of the evaluation. Nevertheless, principals who are engaged with it don't find it to be an easy piece of technology to use as only 53 percent agree that it is relatively easy to use. When asked directly what would make eVAL better, a high school principal replied, “I think one, if it's a little faster. Two, it's kind of difficult to understand at first when you go from putting your notes down and the labeling your notes, to observing, to the reports. There's just so many different parts of this, it's very technical and you have to find things....I don't think it's very user-friendly. The training we got from the ESD wasn't very good, to say the least. In fact, it would be unsatisfactory in many areas.” He clarified, “I think if you have training specific to the framework that you are using, that helps. That would have helped me.” Results from the 2014 survey indicate that 64 percent of principals say they had received professional development to support their use of eVAL.

Interview data from case study districts using eVAL indicate that many principals assumed responsibility for uploading evidence and helping teachers manage the eVAL system. Principals reported a desire to reduce frustration and address time concerns of teachers so that they could focus on more substantive issues in their evaluation. It is important to note that when examining interview data collected later in the implementation process, perceptions of the usefulness of eVAL had changed for a number of principals. In particular, principals note that work done in eVAL across the year resulted in significant time savings when they needed to generate final summative scores and reports for the teachers they were evaluating. Indeed, the survey data bears this out as 57 percent of principals in 2014 indicate that eVAL saves them time in the evaluation process. Some principals credit eVAL with improving dialogue with teachers around evaluations, in terms of having access to the same pieces of evidence, being able to comment on these uploaded artifacts, and being able to send questions to teachers prior to observations.

A superintendent explained that principals who had used the tool reported spending far less time writing comprehensive evaluations than those not using the tool. He indicated, “A lot of principals were taking 8-10 hours, sometimes 12 hours to write a comprehensive evaluation...the ones who were using eVAL said that they could cut that time down significantly because they had entered the stuff and where it all kind of fit, and then they could drop it down into the different components and they had the evidence already there. The ones who were taking 8, 10, 12 hours were going back through everything they had and trying to piece things together and then write the documentation.” Among district administrators surveyed in 2014 who use eVAL, 58 percent feel that eVAL is useful as a platform for sharing information, 68 percent report that it saves them time in the evaluation process, and 90 percent indicate that eVAL is useful in creating a final summary record of the evaluation.

Data from the 2013 and 2014 surveys, along with evidence from interviews with teachers and administrators indicate that during the first two years of implementation, the eVAL tool had both strengths and limitations. Findings support the need for continued professional development to support both teachers and administrators in using technological tools for evaluation purposes. It is important to note that a learning curve exists for educators who are using technological tools, and this can result in frustration and increased workload. However, the increasing use of eVAL, accompanied by continued improvement of the software, has the potential to assist educators in a more efficient and effective implementation of the evaluation process.

Educator Views about the Impact of the Revised Evaluation System

The majority of educators participating in this study agree that the revised evaluation system has a number of benefits, but concerns about time and workload are significant. There is broad agreement that the implementation of the revised system has had a positive impact on professional conversations about what constitutes effective teaching, professional growth of teachers, and high expectations for student learning. On the other hand, nearly all teachers and principals agree that the evaluation system increases their workload, and the vast majority of building and district administrators agree that the biggest obstacle to implementation is time spent on evaluations. Nevertheless, school and district administrators are clear in insisting that they do not want any additional changes to the existing state requirements and policies which could potentially disrupt the work underway. In this section we review the perspectives of educators in the six case study districts.

Educators in the case study districts who participated in the surveys were asked a number of questions about the impact of the revised teacher evaluation system. While it might be expected

that their views would change between the first and second years of statewide implementation, their views on many topics remained remarkably similar.

A majority of teachers either strongly or somewhat agree that the evaluation system will have some benefits. Slightly less than two-thirds of the teachers (62 percent) either somewhat or strongly agree that the evaluation system will prompt them to consider alternative forms of assessment. Additionally, a slight majority either somewhat or strongly agree that the evaluation system will improve their instruction (52 percent) and better align instructional improvement activities in their school or district (56 percent). Teachers also note challenges associated with the revised evaluation system. Nearly all teachers (94 percent) agree that the evaluation system will increase their workload, and 71 percent strongly agree with this. Three-quarters of teachers (76 percent) agree that the evaluation system will primarily be a compliance mechanism of limited professional benefit, with 38 percent strongly agreeing. Table 11 provides details on these items.

Table 11: 2014 Survey: Teacher Views of Evaluation System (n=1823)				
<i>The teacher evaluation system will...</i>	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
improve the quality of my instruction	10%	42%	28%	20%
prompt me to consider alternative forms of assessment	13%	49%	23%	15%
better align instructional improvement activities in my school/district	12%	44%	26%	16%
increase my workload	71%	23%	4%	1%
primarily be a compliance mechanism of limited professional benefit	38%	38%	18%	5%

The 2014 survey data also reveal that a majority of teachers report either a somewhat or very positive impact on the quality of professional collaboration (51 percent). Some teachers report that the opportunities to collaborate have increased as a result of TPEP. As one teacher explained, “I do like all of the feedback that I am getting from my principal and the collaboration opportunities that it gives us.” However, a number of teachers expressed concern about the quality and usefulness of the collaborations around TPEP, and some felt that collaboration actually decreased as a result. According to one teacher, “Teachers are burned out and overburdened by the demands of this new system to the extent that they don't have time to reflect on or develop powerful instruction. I have also witnessed a substantial decrease in collaboration because this system has created a competitive frame of mind wherein teachers feel they need to protect themselves and their ideas.”

Teachers are mixed in how they view the focus on student achievement associated with TPEP. Survey data reveal that a majority of teachers report either a somewhat or very positive impact

on high expectations for student learning (51 percent). Forty-seven percent of teachers indicated a very or somewhat positive impact on student achievement while 44 percent believe it will have no impact (see Table 12). As one teacher explains, “The tracking of student data is not a new thing to me or any of my colleagues. However, now we are required to do it in a particular manner. I do not feel like that is helping the students grow. The reason is because I am spending more time trying to make my data fit into the parameters set and less time actually analyzing and doing interventions to improve the success of my students.” In other words, the prescriptive nature of this work can take away from what teachers actually learn about student growth.

Table 12: 2104 Survey: Perceived Impact of TPEP Implementation (Teacher n=1823, Principals n=142, District Admin n=59)

	Very positive impact			Somewhat positive impact			No impact			Somewhat or very negative impact		
	Teachers	Principals	District Admin	Teachers	Principals	District Admin	Teachers	Principals	District Admin	Teachers	Principals	District Admin
Quality of professional collaboration	9%	20%	29%	42%	66%	56%	35%	11%	7%	14%	4%	7%
High expectations for student learning	11%	20%	27%	40%	59%	53%	40%	18%	15%	8%	3%	3%
Professional growth of teachers	8%	16%	24%	42%	69%	63%	34%	14%	9%	15%	1%	2%
Student achievement	6%	5%	29%	41%	61%	56%	44%	32%	10%	8%	1%	2%

Note: Discrepancies in percentages due to nonresponse

In contrast to teachers’ perceptions, there is broader agreement among school and district leaders that the implementation of the evaluation system has had a positive impact on the quality of professional collaboration, professional growth of teachers, high expectations for student learning and a positive impact on student achievement. Eighty-six percent of building administrators report a very positive or somewhat positive impact of the evaluation system on the quality of professional collaboration, compared with 51 percent of teachers. Similarly, 85 percent of building administrators report a positive impact on professional growth compared with 50 percent of teachers. Other differences can be seen in Table 12.

A majority of principals also agree that the teacher evaluation system has had a positive impact on instructional issues in their building by focusing on issues relevant to their student populations and improving their own understanding of the quality of instruction in the building. It has also prompted a majority of principals to make different staffing decisions and to consider alternative forms of assessment. Table 13 highlights principals’ views on these issues.

Table 13: 2014 Survey: Principal Views on Teacher Evaluation System (n=142)				
<i>The teacher evaluation has...</i>	Strongly agree	Somewhat agree	Somewhat disagree	Strongly disagree
focused on instructional issues relevant to our student population	26%	56%	17%	1%
improved my understanding of the quality of the instruction in my building	42%	44%	11%	2%
prompted me to make different staffing decisions	13%	44%	29%	13%
prompted our school to consider new or alternative forms of student assessment	25%	49%	17%	8%
increased the total number of hours I spend working	78%	16%	6%	1%

The teacher evaluation has increased the number of hours building administrators spend working on evaluations and, in most cases, it has not helped them to be more efficient with their time, nor resulted in the delegation of other aspects of their job. Ninety-three percent of building administrators surveyed agree that the teacher evaluation has increased the total number of hours they spend working. Teachers are also very concerned about the increased workload TPEP has placed upon administrators. The following sentiment was repeated frequently in the open-ended survey responses, “Our administrators are awesome and are 100% behind TPEP, but I feel that it is limiting their accessibility to staff on a daily basis, they have less interaction with students, and their stress levels seem to be at all-time highs.” Teachers reported much gratitude toward their principals and expressed that they felt their principals were their “allies” in this. However, they also worried frequently about the toll that the increased workload was placing upon principals and questioned the efficiency and usefulness of the significant focus on teacher evaluations.

In a similar vein, the biggest obstacle to implementation of the evaluation system according to school and district administrators is time spent on evaluations (66 percent of principals and 63 percent of district administrators considered this a major concern). In addition to these obstacles to implementation, changes to the evaluation system from the legislature or state agencies was identified as a moderate or major concern for 37 percent of building administrators and 46 percent of district administrators. Other concerns of school and district administrators included lack of clarity with regard to expectations, the need for more evaluator training, and perceived authenticity of the evaluation. Overall, confidence in the fairness of the system and opportunities for input regarding implementation were lesser concerns for most administrators (see Table 14).

Table 14: 2014 Survey: Perceived Obstacles to TPEP Implementation (Principals n=142, District Admin n=59)

	Not a concern		Small concern		Moderate concern		Major concern	
	Principals	District Admin	Principals	District Admin	Principals	District Admin	Principals	District Admin
Evaluator training	13%	15%	37%	37%	41%	31%	9%	12%
Confidence in the fairness of the new system	18%	17%	44%	36%	26%	32%	11%	14%
Opportunities for input regarding implementation	23%	20%	38%	46%	26%	22%	11%	9%
Time spent on evaluations	5%	0	12%	3%	17%	32%	66%	63%
Clarity with regard to expectations	7%	9%	40%	37%	34%	34%	18%	17%
Perceived authenticity of the evaluation	15%	9%	37%	41%	31%	37%	15%	12%
Changes to the evaluation system from the legislature or state agencies	6%	0%	18%	15%	37%	34%	37%	46%

Note: Discrepancies in percentages due to nonresponse

For all the challenges in implementing the revised evaluation system, some educators noted that TPEP is one of best models currently available for teacher evaluation, and most have no interest in going back to the old system. A principal explained, “If the purpose of TPEP is to improve instruction, it’s the best model I’ve seen.” School principals in particular are quick to point out that it helps teachers across the spectrum to improve their practice. Several principals pointed out that it has helped their best teachers improve, noting that with a rubric in front of them they know what to do to. Conversations around effective instruction as a result of the instructional improvement focus are frequently mentioned as a tremendous benefit. A superintendent reiterated, “if you talk to anybody who is involved in this, principals or teachers, they would say to you the level of discourse between the administrator and them and the level of discourse in their team around some of this stuff has been significantly deeper and more focused.”

Educators across the system raised concerns about consistency in accountability. A teacher who has been involved in the TPEP rollout from the early days has no personal concerns about how the evaluation will impact her; however, she explained, “My worry as a committee member though is going to be getting everybody on the same page and getting everybody’s questions answered in a clear, concise and thoughtful way that is consistent throughout the district and consistent with the state laws.” A district administrator explained, “... how they are implementing some of the grading things is inconsistent from school to school. So, from a system point of view we should have more district structure, more district systems around some of the common instructional practices so individual principals are not having to do that lift.” Teachers also expressed uncertainty about the process, particularly about what constitutes “evidence” of student growth. As one teacher explained, “the uncertainty of the process is disconcerting. A regular conversation is: 'Would this constitute evidence? Is it powerful enough? Am I on the right track? Is this too much? Is this not enough?'" Teachers also expressed

frustration and concern about the equity of the various student measures used, with a teacher reporting, “I have noticed that teachers have the same evidence but [are] getting different scores on their eval[uation]. Also these same artifacts used for one teacher are not weighed the same as it is for another teacher. What is considered to be a good artifact or piece of evidence?””

The perspectives of educators about the benefits and challenges of the revised evaluation system, together with other data collected from case study districts suggest some specific implications for state and local policy. These implications are discussed in the final section of this report.

Policy Implications and Conclusions

This study follows the early TPEP implementation efforts of six districts in Washington state. It should be noted that these districts are not representative of districts statewide. As early implementers of TPEP, their selection for the study was deliberate to provide a window into the issues and challenges faced by districts that have expended tremendous effort, thought and care in doing this work. These districts embraced the idea of implementing TPEP in an authentic way. While each of the districts has experienced its own challenges in the early years, they reflect in some ways, best case scenarios. In this last section of the report, we focus on conclusions and potential policy implications based on what we have learned from the experiences of these educators.

Summary and Concluding Comments

The fundamental tension between evaluation for accountability and evaluation for continuous improvement appears to be present in the case study districts. In these districts, there is evidence that leaders are paying attention to the kinds of messages they send to school-based staff regarding the connection between the district’s goals for teaching and learning and the revised evaluation system. While district leaders (including teachers, union leaders and building leaders involved in TPEP efforts) recognize that state policies require fundamental change, they attempt to frame the conversation about the changes in ways that could potentially contribute to the improvement of teaching and learning in the district.

One superintendent explained the district’s efforts to leverage the evaluation for instructional improvement this way: “I don’t want to play around the edges of this. I don’t want to create things that say OK we showed student growth. I want genuine student growth, so if we are going to do this and put our time and energy into this, I think this has the possibility of really changing the face of what goes on in the classroom....” But he also acknowledged the potential downside, “I think there is a real way to play the game and this could be a lot of energy that doesn’t get us the kind of results I hope we get....”

The design of state policy did allow for local decision-making discretion about several aspects of the revised system, thereby increasing the potential likelihood that districts would experience more ownership and potential positive impact from the initiative. Given this local decision-making discretion, it is not surprising that variation exists in how districts have responded. Districts vary in their choice of instructional framework and their implementation schedule. Teachers vary in their choice of goals, the ways decisions about goal setting are made and perhaps more significantly, in the variety of forms of evidence of student growth they are using. Districts also vary in the type and amount of professional learning opportunities provided for teachers to learn about the revised evaluation systems. In particular, school and district leaders indicate that it has created space for some of the best professional conversations they've ever had with their staff about what good teaching looks like.

There are also initial indicators of areas for concern. Nearly all educators report a significant increase in their workload that they attribute to involvement in the revised system, and administrators are worried about whether state policy changes will occur that may disrupt the work that is currently taking place. There is also evidence of disconnections between the views of teachers and administrators on certain aspects of the evaluation system, such as staff capacity to set appropriate student learning goals and collect evidence that can be used for these purposes, as well as the capacity of building administrators to lead these efforts. Finally, a lack of assessment tools that are relevant and useful in gauging student growth may detract from effective implementation and undermine confidence in the revised system.

Potential Policy Implications

Finally, we present potential policy implications that may help inform school and district leaders, and state policymakers, as they consider how to support and sustain the efforts of schools and districts to productively engage staff in the revised evaluation system.

Allow time for implementation of complex policy

TPEP has resulted in a massive and complicated change in school policy and practice in Washington state. The revised evaluation system requires a reconceptualization of the role of the principal and necessitates that school leaders develop deep expertise in instruction and spend substantial portions of their time observing and providing feedback to teachers. Educators in the case study districts took the implementation effort very seriously, but changes of this magnitude take years, as they touch many aspects of schooling. A district administrator explained, "Change takes time. It's a five to seven year process... I've been in education, I don't know, 25 years or something, and I can't ever remember when we completely changed the teacher evaluation

system, the principal evaluation system, and all the standards we are teaching to at the same time.”

Stay the course – no more policy changes

Many school and district leaders consider TPEP to be the best option available at the present time for teacher evaluation, and they would encourage state policymakers to stay the course. The biggest obstacle to implementation is time spent on evaluations. School and district administrators are clear in insisting that they do not want any additional changes to the existing state requirements and policies which could potentially disrupt the work underway. Educators indicate that they are becoming more comfortable with the process, and developing more trust in the system as teachers work through it. Most would also like to stay the course on student growth using formative assessments. Regarding formative assessments, a district administrator explained, “We know that the most powerful assessments are the ones that reflect what happened in that classroom each day.”

Invest in capacity building and coherence

The case study districts found that additional administrators and restructuring of human resources were needed to conduct and support the evaluation process. Districts can also seek ways to connect the evaluation system to an instructional improvement agenda that supports a conception of teaching and learning for continuous improvement. In this way, districts can bring coherence to the many aspects of system change that has been prompted by TPEP. Educators should seek to capitalize on the professional conversations with teachers about the progress of their students.

Provide for continued support for integrated professional development

As the state offers opportunities for TPEP training, they should be designed in such a way as to be easily integrated with other pressing needs such as Common Core State Standards, differentiated instruction for subpopulations of students, and alternative assessments. Districts indicate a need for additional, appropriate assessments, which can support the integrity and reliability of the evaluation process, as well as training in how to use assessments to inform student growth goals. Support for and refinement of electronic tools that can support the evaluation process are also needed.

Engage support of teacher preparation institutions

The TPEP model has many elements in common with the content provided as part of the preparation of beginning teachers. Educators have noted that in some ways, TPEP implementation is easiest for novice teachers who have been trained in areas such as classroom assessment, use of instructional frameworks, goal setting, and evidence collection as part of their preparation program.

References

- Bill and Melinda Gates Foundation (2013). *Ensuring fair and reliable measures of effective teaching*. Seattle, WA: Author.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. 2nd Edition. Thousand Oaks CA: Sage.
- Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership* 68 (4): 35-39.
- Darling-Hammond, L. Amrein-Beardsley, A., Haertel, E. & Rothstein, J. (2012). Evaluating teacher evaluation. *Kappan* 93 (6): 8-15.
- Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement*. New York: Teachers College Press.
- Goe, L., Biggers, K., & Croft, A. (2012). *Linking teacher evaluation to professional development: Focusing on improving teaching and learning*. National Comprehensive Center for Teacher Quality. Washington, DC: Author.
- Goe, L. & Holdheide, L. (2011). *Measuring teachers' contributions to student learning growth for nontested grades and subjects*. National Comprehensive Center for Teacher Quality. Washington, DC: Author.
- Harris, D. (2011). *Value-added measures in education: What every educator needs to know*. Boston, MA: Harvard Educational Press.
- Harris, D. (2009). Would accountability based on teacher value-added be smart policy? An examination of the statistical properties and policy alternatives. *Education Finance and Policy* 4 (4): 319-350.
- Herlihy, C., Karger, E., Pollard, C., Hill, H., Kraft, M., Williams, M., & Howard, S. (2014). State and local efforts to investigate the validity and reliability of scores from teacher evaluation systems. *Teachers College Record* 116 (1): p.- <http://www.tcrecord.org/library> ID Number: 17292.
- Isore, M. (2009). *Teacher evaluation: Current practices in OECD countries and a literature review*. OECD Education Working Paper. Paris: OECD.

Kersting, N., Chen, M., & Stigler, J. (2012). Value-added teacher estimates as part of teacher evaluations: Exploring the effects of data and model specifications on the stability of teacher value-added scores. *Educational Policy Analysis Archives* 21 (7).
<http://epaa.asu.edu/ojs/article/view/1167>.

Humphrey, D., Koppich, J., Bland, J., & Bosetti, K. (2011). Peer review: Getting serious about teacher support and evaluation. Menlo Park, CA: SRI International

Lavigne, A. (2014). Exploring the intended and unintended consequences of high-stakes teacher evaluation on schools, teachers, and students. *Teachers College Record* 116 (1):
<http://www.tcrecord.org/library> ID Number: 17294.

Looney, L. (2011). Developing high-quality teachers: Teacher evaluation for improvement. *European Journal of Education* 46 (4): 440-455.

McCaffrey, D., Lockwood, J., Koretz, D., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy* 4 (4): 572-606.

Munson, B. (1998). Peers observing peers: The better way to observe teachers. *Contemporary Education* 69: 108-110.

OECD (2005). *Teachers matter: Attracting, developing and retaining effective teachers*. Paris: Author.

OECD (2009). *Creating effective teaching and learning environments: First results from TALIS*. Paris: Author.

Papay, J. (2012) Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review* 82 (1): 123-141.

Reardon, S. & Raudenbush, S. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy* 4 (4): 492-519.

Taylor, E. & Tyler, J. (2011). *The effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers*. NBER Working Paper 16877.

Toch, T. & Rothman, R. (2008). *Rush to judgment: Teacher evaluation in public education*. Washington, DC: Education Sector.

Steele, J., Hamilton, L. & Stecher, B. (2010). *Incorporating student performance measures into teacher evaluation systems*. Santa Monica, CA: RAND Corporation.

Weisberg, D. Sexton, S., Mulhern, J. & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. The New Teacher Project. Brooklyn, NY: Author.